DOCUMENT RESUME

ED 314 469 TM 014 350

AUTHOR	Ziomek, Robert, Ed.; Wolmut, Peter, Ed.
TITLE	National Association of Test Directors 1989 Symposia, Paper, and Survey. Papers presented at the Annual Meeting of the American Educational Research Association (San Francisco, California, March 27-31, 1989).
INSTITUTION	National Association of Test Directors, Cedar Rapids, IA.
PUB DATE	Mar 89
NOTE	301p.
PUB TYPE	Collected Works - Conference Proceedings (021) -- Tests/Evaluation Instruments (160)
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	*Academic Achievement; Adolescents; Children; Day Care; Dropout Prevention; Educational Assessment; Educational Research; Elementary Secondary Education; Evaluation Methods; *Evaluation Utilization; Magnet Schools; Measurement Techniques; National Surveys; School Districts; *School Effectiveness; *Standardized Tests; Test Use
IDENTIFIERS	*Educational Indicators; *National Association of Test Directors

ABSTRACT

The papers in this collection reflect topics of major interest to members of the measurement and testing communities. Symposium 1, "Beyond the Wall Chart: Indicators and School Ranking," includes the following papers: (1) "What Our 17-Year-Olds Know: One District's Assessment" (L. D. Wilkinson and D. L. Galindo); (2) "Indicators of Equity and Excellence: Searching for Consensus" (L. H. Brown and J. Litzenberger); (3) "Development of a 50-State System of Educational Indicators: Issues of Design, Implementation, and Use" (R. K. Blank); (4) "Research, National Assessment, and School Evaluation in the U.K." (C. Gipps); (5) "The Ranking of Schools and the Florida Norm-Referenced Testing Program" (M. Baron); (6) "Ranking School Districts on the Basis of Statewide Results: Is It Meaningful or Misleading?" (T. R. Guskey); (7) "Discussion I" (R. Ziomek); (8) "Discussion II: The Name of the Game" (P. Wolmut); and (9) "Discussion III: Comments to Public School Researchers, Evaluators, and Testing Practitioners" (G. Ligon). A second symposium, "Dog-Eared Reports Are the Best of Breed: Recognizing and Rewarding Evaluation Utilization" contains the following papers: (1) "Schenley High School Teacher Center" (P. LeMahieu); (2) "Eight Individual Magnet Schools" (E. Tomblin); (3) "Elementary Language Arts" (L. Baldwin); (4) "TEAMS Mastery and District Dropout Prevention Effectiveness" (G. Ligon); and (5) "School-Age Child Care" (S. Mitchell). The paper "Standardized Tests at the Pre-School Level: Do They Tell Us the Truth about Children?" (E. Pardo and J. C. Russell) is included, as well as the survey report "Testing Issues and Practices" (E. Bauer). (SLD)

ED314469

NATIONAL ASSOCIATION OF TEST DIRECTORS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
● Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Peter Wolmut
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

1989 SYMPOSIA,

I. Beyond the Wall Chart: Indicators and School Ranking

[illegible]

TM 014 350

NATIONAL ASSOCIATION OF TEST DIRECTORS

We are pleased to present this, the fifth volume of published symposia, papers, and surveys of the National Association of Test Directors. The Association considers the promotion of intellectual discourse on testing matters (from both theoretical and practical perspectives) to be essential to its professional mission. The publication of this document is one important activity that is undertaken to address that goal. The papers contained herein were first presented at the March, 1989 meeting of the American Educational Research Association in San Francisco, CA. They reflect topics of major interest to members of the measurement and testing communities.

Robert Ziomek
CEDAR RAPIDS PUBLIC SCHOOLS

Peter Wolmut
MULTNOMAH (OR) EDUCATION SERVICE DISTRICT

TABLE OF CONTENTS

Symposium I
Beyond the Wall Chart: Indicators and School Ranking

Introduction.. 4

What Our 17-Year-Olds Know: One District's Assessment.......... 5
L. David Wilkinson
D. Letticia Galindo

Indicators of Equity and Excellence: Searching for Consensus.... 62
Lee Hillman Brown
Jerry Litzenberger

Development of a 50-State System of Educational Indicators: Issues of Design, Implementation, and Use...................... 97
Rolf K. Blank

Research, National Assessment, and School Evaluation in the UK..127
Caroline Gipps

The Ranking of Schools and the Florida Norm-Referenced Testing Program................154
Marc Baron

Ranking School Districts on the Basis of Statewide Results: Is It Meaningful or Misleading?................................159
Thomas R. Guskey

Discussion I...182
Robert Ziomek

Discussion II: The Name of the Game............................184
Peter Wolmut

Discussion III: Comments to Public School Researchers, Evaluators, and Testing Practitioners...........................188
Glynn Ligon

(Continued Next Page)

TABLE OF CONTENTS
(Continued)

Symposium
Dog-Eared Reports are the Best of Breed: Recognizing and Rewarding Evaluation Utilization

Introduction..190

Schenley High School Teacher Center........................191
Paul LeMahieu

Eight Individual Magnet Schools..........................214
Elizabeth Tomblin

Elementary Language Arts..............................219
Lee Baldwin

TEAMS Mastery and
District Dropout Prevention Effectiveness....................230
Glynn Ligon

School-Aged Child Care...............................237
Stephanie Mitchell

Paper

Introduction..250

Standardized Tests at the Pre-School Level:
Do They Tell Us the Truth About Children?....................251
Elly Pardo
Joanne Collins Russell

Survey

Introduction..283

Testing Issues and Practices............................284
Ernest Bauer

THE VIEWS EXPRESSED HEREIN ARE THOSE OF THE RESPECTIVE AUTHORS AND NOT NECESSARILY THOSE OF THE NATIONAL ASSOCIATION OF TEST DIRECTORS OR ITS BOARD OF DIRECTORS.

AUTHORS AND EDITORS

LEE BALDWIN
Director of Testing/Prog. Eval.
Orange County Public Schools
PO Box 271
Orlando FL 32802

MARC BARON
Director of Testing
Palm Beach County Schools
3323 Belvedere Road
West Palm Beach FL 33402

ERNEST BAUER
Director of Testing
Oakland Public Schools
2100 Pontiac Lake Rd.
Pontiac MI 48054

ROLF K. BLANK
Project Director
Council of Chief State School Officer
400 N. Capitol St., Suite 379
Washington DC 20001

LEE HILLMAN BROWN
Director of Testing
Edmonds Public Schools
12214 SE 18th Pl.
Bellevue WA 98005

D. LETTICIA GALINDO
Evaluation Associate
Austin ISD
6100 Guadalupe
Austin TX 78752

CAROLINE GIPPS
Curriculum Studies Dept.
London Institute of Education
20 Bedford Way
London WC1H OAL U.K.

THOMAS R. GUSKEY
Professor of Education
University of Kentucky
131 Taylor Education Bldg.
Lexington KY 40506-0001

PAUL LEMAHIEU
Director, Research/Testing/Eval.
Pittsburgh Public Schools
341 S. Bellefield Ave.
Pittsburgh PA 15213

JERRY LITZENBERGER
Director, Research/Evaluation
Bellevue Public Schools
12111 NE 1st St.
Bellevue WA 98009

GLYNN LIGON
Director of Management Info.
Austin ISD
6100 Guadalupe
Austin TX 78752-4495

STEPHANIE MITCHELL
Evaluation Specialist
Portland Public Schools
P. O. Box 3107
Portland OR 97208-3107

ELLY PARDO
Senior Program Director
Boston Public Schools
26 Court St.
Boston MA 02108

JOANNE COLLINS RUSSELL
Program Director, E. L. C.
Boston Public Schools
50 Beechcroft St.
Brighton MA 02135

ELIZABETH TOMBLIN
Director of Evaluation
San Diego City Schools
4100 Normal St.
San Diego CA 92103-2682

L. DAVID WILKINSON
Evaluator
Austin ISD
9710 Moorberry St.
Austin TX 78729

PETER WOLMUT
Director, School Support Services
Multnomah ESD
PO Box 16657
Portland OR 97216

ROBERT ZIOMEK
Director, Research/Planning/Testing
Cedar Rapids School District
346 Second Ave., S.W.
Cedar Rapids IA 52404

ACKNOWLEDGEMENTS

The editors wish to express deepest appreciation to the members of the Board of the National Council on Measurement in Education and particularly Janie Hall, 1989 Program Chair of Division H, American Educational Research Association, for their continued support of National Association of Test Directors endeavors. Special thanks go to Juanita Dickson and Charlene Smith for their assistance in producing this volume.

Symposium I

BEYOND THE WALL-CHART: INDICATORS AND SCHOOL RANKINGS

Local district test directors find that the annual reporting of rankings by state on the Secretary of Education's "wall chart", along with an SAT "bravo" or fingershake leaves much to be desired in terms of what they are trying to report to their communities about local test results. In the merger of a pair of symposia organized by Robert Ziomek and Janie Hall, David Wilkinson and Letticia Galindo describe their study of Austin student performance on the same Literature and History items which were the basis for national criticism by Ravitch and Finn. Lee Brown and Jerry Litzenberger present a set of indicators developed by a consortium of 14 Washington State suburban school districts, consequent to those districts having decided to report disaggregated achievement test results in the same fashion. The Council of Chief State School Officers continues to support the concept of providing indicators which will effectively rank the 50 states. Rolf Blank reviews the general principles which CCSSO has established, followed by examples of how these have been applied to the definition of science indicators. Caroline Gipps outlines the parallel national system of school accountability which has been developed in the United Kingdom. Marc Baron describes the process and implications of the ranking of Florida schools on their state-wide norm-referenced program, while Thomas Guskey examines Kentucky's system of ranking their school districts, noting some of the problems found there. Robert Ziomek, Peter Wolmut, and Glynn Ligon present different summative perspectives of where we are all going with this and what it means.

WHAT OUR 17-YEAR-OLDS KNOW: ONE DISTRICT′S ASSESSMENT*

L. David Wilkinson
D. Letticia Galindo
Austin Independent School District

Introduction

This paper presents the results of the only assessment (that we know of) by a local school district of its 17-year-olds' knowledge of history and literature using the same test items administered by the National Assessment of Educational Progress (NAEP). The results of the national assessment conducted in spring, 1986, using a sample of 7,812 students were much publicized, and were even the subject of a best-selling book, What Do Our 17-Year-Olds Know?, by Diane Ravitch and Chester E. Finn, Jr. (Harper & Row, copyright 1987). In this book, the authors deplored the "serious gaps" they found in students' basic knowledge of history and literature and issued a call for improving teaching and learning in these areas.

In the Austin Independent School District (AISD), Austin, Texas, the national furor over the apparently deplorable state of students' knowledge of history and literature became infused with an ongoing discussion about the amount of content being taught in AISD classrooms and the appropriate balance between skills and knowledge. Highlighted by the national debate, these concerns became a focus of attention by the Superintendent and administration in the 1987-88 school year.

The question was raised, "How does the District determine what the content of instruction ought to be?" Taking this question as a point of departure, the Office of Research and Evaluation (ORE) undertook a pilot study of the content in literature and American history curricula. The results of this study are the topic of this paper.

Objectives of This Paper

This paper will address four objectives:

1. To describe a pilot study of the content of the American history and literature curricula in an urban school district,

2. To present results from the District's administration of the NAEP items compared with the national results as reported in What Do Our 17-Year-Olds Know?, and results from local review of NAEP item specifications,

---

*The opinions and conclusions expressed herein are those of the authors and do not necessarily reflect the position or policy of the Austin Independent School District or the Office of Research and Evaluation. No official endorsement should be inferred.

3. To discuss problems encountered in carrying out the study, particularly local campus resistance, and

4. To discuss directions for future study and implications for other school districts.

Data Sources

The results reported in this paper derive from two sources: (1) the District's spring, 1988, administration to 960 high school students of the 141 U.S. history and 121 literature items used by the National Assessment of Educational Progress (NAEP) in spring, 1986, and (2) the rankings of NAEP item specifications for U.S. history and literature items by local reviewers. To honor an agreement with NAEP to maintain the security of its materials for future research, actual test items are not disclosed. District results are compared with the national results using the same, disguised item descriptors used in *What Do Our 17-Year-Olds Know?*. Individual responses from reviewers are also protected.

AISD's Study

Origin of the Study

In the fall of 1987, AISD's Superintendent broached the idea of administering the NAEP history and literature tests to some AISD students. At the time, ORE was in the process of developing its research and evaluation agenda for 1987-88. The Superintendent's interest was therefore construed as a request for an addition to the agenda, and action was taken accordingly. An evaluation plan was drawn up which placed the testing in a framework of an ongoing evaluation of the District's curricula, beginning with the content of the curriculum in U.S. history and world literature courses. The selection of these curricular areas for study was justified by reference to the controversy about instruction in American history and literature engendered by the NAEP administration and the publication of results in *What Do Our 17-Year-Olds Know?* Subsequent years' evaluations, to be based on the first-year study, were to be concerned with other curricular areas. This ambitious plan elicited favorable comment from the Superintendent and was adopted in November, 1987, as part of the official research and evaluation agenda for 1987-88 by AISD's Board of Trustees.

Study Questions

Four fundamental questions served as the framework around which the study of the content in the American history and literature curricula was organized:

1. What are the important things that we should teach our children?

2. Are the important things in our curriculum?

3. Are we teaching the important things to our children?

4. Are children learning the important things?

Evaluation Activities

Evaluation activities were planned to address each of these questions. In the order of the questions, these were:

1. To have a group of persons with a stake in instruction in the District (teachers, instructional coordinators, subject area professors from area universities, etc.) review the item specifications from the National Assessment of Educational Progress and select those of greatest importance for the students of the District;

2. To have instructional coordinators from elementary and secondary education review the Texas essential elements (mandated curriculum), the scope and sequence of textbooks, and curriculum guides for evidence that the important content is included in the District's planned curriculum;

3. To collect teachers' tests for analysis by instructional coordinators to ascertain what is being taught (on the assumption that what is being tested is being taught); and,

4. To administer those items from the National Assessment earlier identified as measuring the important content in U.S. history and literature to the District's eleventh graders so that their performance could be compared with the performance of students nationwide.

At the conclusion of the 1987-88 school year, only the first and fourth of these activities were substantially completed. A sample of teachers' tests was collected, but their content has not been analyzed. The proposed review of the District's planned curriculum has not been implemented as of this writing. To date, owing to budget cuts and other administrative priorities, there is no plan to revive or continue the study.

History and Literature Testing

Although the student testing was originally conceived as occurring after review of the NAEP item specifications, as concerns about the testing delayed getting the study underway and other activities encroached on the evaluation timeline, it became apparent that the testing had to proceed or risk not being conducted in the 1987-88 school year. Therefore, although it was logically out of sequence with the review of item specifications, it was decided to administer all of the history and literature items to the District's eleventh graders, rather than a subset of items judged by local reviewers to be most applicable to Austin students, as originally planned. It was determined that the review of item specifications would occur either concurrent with or after the testing.

Test Sample

It was originally proposed that the history and literature tests be given to all of the District's eleventh graders. However, input from high school principals and administrators in Secondary Education prompted a revised plan in which one half of the District's eleventh graders would take the test, while the remainder of the students took the Social Studies Test from the Tests of Achievement and Proficiency (TAP), an annual achievement test normally given to all high school students in the spring. An alternate, but not recommended, proposal had the students taking the history and literature tests in addition to the TAP.

In a meeting with secondary principals held February 10, 1988, the ground rules for the testing were finally established. The Superintendent was quite specific in his instructions to ORE:

1. The testing of eleventh graders in history and literature would be voluntary for both schools and teachers, rather than mandatory as originally proposed.
2. A sample of students would be tested, but random selection would not be a concern.
3. Testing would be an in-house exercise conducted to obtain some information for examining the District's curriculum and to promote discussion about how much content was being taught and how much students were learning.
4. The results were not to be published as an "official" ORE study, nor be released to the media.

A memo was sent on February 29, 1988, to the 10 high school principals asking them to designate a contact person with whom ORE could work out the details. Also, they were to send to ORE a detachable section of the memo documenting the number of students to be tested in history and/or literature and the date(s) of test administration. Follow-up calls were made by ORE staff to those campuses which did not submit written documentation.

Figure 1 depicts the participating schools, the number of tests administered in each subject area, and the test dates. Figure 2 tables the number and percent of tests administered in each area by the grade level, sex, age, and ethnicity of the students tested. The numbers of students taking one or the other or both tests are shown in Figure 3.

Seven of the 10 AISD high school campuses administered a total of 1,377 history and literature tests, 695 in history and 682 in literature. A total of 960 students were tested, with 4.3.4 . Most, but not all, of the students tested were eleventh graders (89.4% in history and 91.6% in literature). Most of the students tested were either 16 or 17 at the time of the testing. (Age was as of April 22, 1988, the date roughly at the midpoint of the time span covered by the testing.) Approximately equal percentages of boys and girls were tested. Almost two thirds of the students tested were White/Other. (Small percentages of Asian and Native American students were included in this category.) Approximately equal percentages of Black and Hispanic students were tested.

Figure 1. FREQUENCY COUNT OF TESTS ADMINISTERED BY SCHOOL IN THE AISD HISTORY AND LITERATURE TESTING, SPRING, 1988

SCHOOL	TEST	NUMBER OF STUDENTS	TEST DATE(S)
Austin	History	137	April 11
Crockett	History	111	April 22
	Literature	110	April 22
Johnston	History	60	April 5
	Literature	191	April 11-15
LBJ	History	95	May 12
	Literature	99	May 12
McCallum	History	253	May 12
	Literature	245	May 12
Reagan	History	37	May 11
	Literature	37	May 11
Robbins	History	2	April 4-8
	Literature	0	April 4-8
TOTAL	History	695	
	Literature	682	
		1,377	

Figure 2. FREQUENCY COUNTS OF HISTORY AND LITERATURE TESTS ADMINISTERED, BY GRADE, SEX, AGE, AND ETHNICITY

	HISTORY		LITERATURE	
	Number	Percent	Number	Percent
Grade				
Unknown	3	.4%	0	.0%
9	2	.3%	1	.1%
10	32	4.6%	33	4.8%
11	621	89.4%	625	91.6%
12	37	5.3%	23	3.4%
	695	100.0%	682	100.0%
Sex				
Male	351	50.5%	347	50.9%
Female	344	49.5%	335	49.1%
	695	100.0%	682	100.0%
Age				
14	1	0.1%	0	0.0%
15	6	0.9%	5	0.7%
16	249	35.8%	260	38.1%
17	369	53.1%	370	54.3%
18	61	8.8%	40	5.9%
19	7	1.0%	6	0.9%
20	2	0.3%	0	0.0%
21	0	0.0%	1	0.1%
	695	100.0%	682	100.0%
Ethnicity				
White/Other	442	63.6%	441	64.7%
Black	129	18.6%	120	17.6%
Hispanic	124	17.8%	121	17.7%
	695	100.0%	682	100.0%

Figure 3. FREQUENCY COUNTS OF AISD STUDENTS TAKING THE HISTORY AND LITERATURE TESTS

	Number	Percent
History Only	278	29.0%
Literature Only	265	27.6%
Both Tests	417	43.4%
	960	100.0%

Test Items

AISD used the full complement of NAEP items--262 knowledge questions, 141 in history and 121 in literature. As received from NAEP, these questions were divided up among four books in each area, labeled: H1, H2, H3, H4 and L1, L2, L3, and L4. Each student tested received from 34-36 history questions and from 30-31 literature questions, depending on which booklets the student received.

Test Time

Total time for a student administered both the U.S. history and literature tests was approximately 40 minutes.

Testing Procedures and Materials

During March and April, 1988, ORE staff drafted and prepared materials to be disseminated to the principal, the Building Test Coordinator (BTC), teachers, and students prior to testing. This information's purpose was threefold:

1. To obtain an agreement of security and confidentiality regarding the test items,

2. To instruct select personnel on testing procedures so as to facilitate the process, and

3. To provide basic information on the test and its objectives.

Listed below, and included as Attachments 1 through 5 are:

- Oath of Security and Confidentiality
- BTC Checklist
- Teacher Checklist
- Teacher Fact Sheet
- Student Fact Sheet

These materials were sent to the campuses along with the designated number of presorted history and/or literature booklets and answer sheets. A standard, mark-sense answer document was adapted for use as an answer sheet. Where teachers names were furnished to ORE, student answer sheets were preslugged with student identifying information from central computer files. Attachments 6 and 7 are samples of preslugged history and literature answer sheets.

Test Processing

Student answer sheets were scanned by District personnel using the District's IBM optical scanner. Computer programs were written by an ORE programmer to score the tests based on right/wrong information supplied by NAEP.

Test Results

Test results were produced after the last tests were administered on May 12, 1988. Each campus received schoolwide results by test area with items arranged in descending order according to the percentages of students nationwide who answered the item correctly. The number of students receiving the item, the number answering correctly, and the percentage of students answering it correctly were printed for each item. No individual or class results were created. Answer sheets returned without form numbers indicating the booklet number (e.g., H1, L2, etc.) could not be scored and were therefore excluded from the results. As a result, one school did not receive its test results.

Representativeness of the AISD Sample

A caveat is offered before proceeding with the test results. It is not clear to what extent the students chosen by the schools to be tested were representative of AISD eleventh-grade students overall or how comparable the AISD students who were tested were to the students tested in the national assessment. As Figure 1 shows, a varying number of students were tested on each campus. One of the schools administered only the history test. One high school with 291 eleventh-grade students only administered the tests to about one class. Other high schools tested most of their eleventh graders, e.g., McCallum High School. Three high schools, with a total of 1,098 eleventh graders, did not participate in the testing, however.

Figure 2 indicates that not all of the students tested were eleventh graders, nor were all of the students 17 years old. As Figure 3 portrays, some students took only one of the tests, while other students took both tests.

In terms of their demographic characteristics, the AISD students who were tested (mostly eleventh graders) were similar to AISD eleventh graders at the time of the testing. There were approximately equal percentages of males and females in both the test sample and the general population. Most of the students tested (90.7%) and most of the AISD eleventh graders (84.8%) were 16 or 17 years old. There were slightly greater percentages of White/Other students (+6.8%) and slightly fewer Hispanic students (-5.2%) in the test sample than in the population of eleventh graders overall.

In comparison with the national sample, some AISD students were older and some younger than the 17-year-olds tested nationally. Some AISD students were in lower grades than grade 11, and some were in grade 12. Distribution by gender was about the same. See Figure 2. However, AISD tested a smaller percentage of White/Other students than was tested nationally, 64.1 to 82.2, and a larger percentage both of

Black and Hispanic students--18.1% versus 12.9% and 17.8% versus 5.9%, respectively. If lower achievement is associated with the presence of minority test takers, then other things being equal, the greater representation of minority students in AISD's sample than in the national sample would argue that the performance of AISD students should be worse than that nationally. This was not the case as will be discussed in the next section.

While it is not possible to say with any certainty that the AISD students tested were representative of eleventh graders as a whole, the test sample did not appear to be greatly different from the general population of AISD eleventh graders, at least in terms of demographic characteristics. The nonparticipating high schools did not have materially different student bodies than the participating schools, nor, despite the researchers' fears, were there any indications that local campus personnel selected solely honors students or the like to be tested. Therefore, it may be reasonable to assume that the AISD students tested, while not randomly selected, may have been fairly representative of AISD eleventh graders.

Comparability with the national sample is less clear. The demographic comparison with the national sample has already been discussed. Special education students were probably excluded in the AISD administration, but NAEP guidelines afforded ample opportunity for schools to exclude special education students from the national administration. In the future, if resources become available, it will be possible to examine the AISD sample to determine if it is representative and comparable to the national sample.

With these cautions and qualifications expressed, it may nonetheless be instructive to form some impressions of the results.

What AISD Students Know

Comparison of AISD and National Results

AISD's results are presented in Attachment 8. Attachment 8 is a rank-order listing by test area of AISD results arranged in descending order according to the percentages of students nationwide who answered the item correctly. The national results are reproduced in Attachment 9.

AISD Percent Correct and National Percent Correct

There are many ways in which the results could be analyzed. One way is by comparing the AISD's percentage correct on each item with the percentage correct obtained by students nationally. For comparison purposes, we adopt as a standard the scale used by Ravitch and Finn in which a score of 100 is "perfect," scores between 80 and 100 are treated as if they were A's and B's, scores between 60 and 80 as C's and D's, and below 60 is taken as a failing mark.

History. On their scale, Ravitch and Finn found among students nationwide that 15 questions were answered correctly by 80% or more of the students, an additional 45 questions were answered correctly by from 60% to 80% of the students, and 81 questions were answered correctly by fewer than 60% of the students. This means that 60 of 141 questions (42.6%) were "passed" (answered correctly by at least 60% of the students), while 81 questions (57.4%) were "failed."

By comparison, in AISD, 29 questions were answered correctly by 80% or more of the students, 57 additional questions were answered correctly by from 60% to 80% of the students, and 55 questions were answered correctly by fewer than 60% of the students. In other words, 86 of 141 questions (61.0%) were "passed," while 55 questions (39.0%) were "failed."

On what kinds of history questions did AISD students do best and worst? A great deal of attention was paid to the inability of most students to place the Civil War in the correct half-century. In the national sample, only 32.2% of the students could do so correctly; in AISD, only a slightly greater 40.1% answered the question correctly. The lowest score dealing with chronology obtained by the students tested nationally in 1986 involved placing Abraham Lincoln's presidential term in its proper twenty-year period (1860 to 1880). Only 24.7% of the students tested nationally made the correct placement; even fewer AISD students (20.2%) could do so two years later in 1988.

In the knowledge area of maps and geography, 92.1% of the students in the national sample could identify the Soviet Union on a map of Europe; 96.5% of AISD students could do so (the highest percentage correct for any history question). On questions relating to the U.S. Constitution and its antecedents, 30.6% of the students in the national sample correctly identified the Magna Carta, while 33.5% of AISD students did so. Regarding women in history, 83.8% of students nationwide knew that Harriet Tubman helped slaves escape; a similar 86.6% of AISD students knew of her role. On questions dealing with more recent events, Sputnik was recognized by 62.7% of the students in the national sample and by 77.3% of the AISD students tested. Only 53.9% of students nationally could correctly depict the American role in Korea, compared with 66.5% of AISD students. In the area of civil rights, only 55.5% of students nationally accurately described the Brown decision. An "adequate" (to use Ravitch's and Finn's term) 74.1% of AISD students recognized its significance.

Literature. On the literature assessment, nationally, 14 of the 121 questions were answered correctly by 80% or more of the students, 37 more questions were answered correctly by from 60% to 80% of the students, and 70 questions were answered correctly by fewer than 60% of the students. Using the same scale as for history, this means that *51 of 121 questions (42.1%) were "passed"* (answered correctly by at least 60% of the students), *while 70 questions (57.9%) were "failed."*

By comparison, in AISD, 22 questions were answered correctly by 80% or more of the students, 45 additional questions were answered correctly by from 60% to 80% of the students, and 54 questions were answered correctly by fewer than 60% of the students. In other words, *67 of 121 questions (55.4%) were "passed," while 54 questions (44.6%) were "failed."*

Comparing individual items, among students nationwide, less than 20 percent could associate each of these authors with their most prominent works: de Tocqueville (15.5%), Joyce (15.6%), Dostoevsky (17.0%), Ellison (18.3%), and Conrad (19.3%). Recognition among AISD students of the same authors and their works was 26.4%, 17.9%, 30.9%, 30.9%, and 26.1%, respectively. Nationally, 57.0% of the students knew that Shakespeare wrote sonnets as well as plays; 76.5% of AISD students knew this fact.

Among questions related to the Bible, 61.0% of students in the national sample and 60.4% of AISD eleventh graders knew that King Solomon was renowned for his wisdom. On questions about Greek mythology, 45.4% of students nationwide correctly attributed the start of the Trojan War to the abduction of Helen; 54.5% of AISD students identified the cause of the war. On questions related to epic literature, only 36.1% of the 17-year-olds in the national sample identified Geoffrey Chaucer as the author of *The Canterbury Tales*. A still-unacceptable 45.8% of AISD students identified the author.

On items concerning well-known passages and quotations, 88.1% of students in the nation recognized Martin Luther King, Jr.'s "I have a dream," the fourth highest score on the entire literature assessment. The passage elicited the highest percent correct score among AISD students (95.0%) of all 121 literature items.

The items answered by the smallest percentages of students nationwide and in AISD concerned Blake's image of a tiger in his famous poem and the theme of Bunyan's *Pilgrim's Progress*. Nationally, only 13.6% of the students answered the Blake item correctly, compared with 13.2% of AISD students. Only 13.4% of students nationwide and an equally poor 13.5% in AISD knew what Bunyan's novel was about.

Performance at the Top and Bottom Quarters

Another approach for comparing AISD to the nation is to compare the performance of AISD high schools to the national sample on the items in the top and bottom quarters based on national performance. In history, this would include 35 items at the top and another 35 at the bottom (of 141 items). In literature, this would include 30 items at the top and 30 at the bottom (of 121 items).

History. In the top quarter, AISD's performance exceeded that of the nation on 20 of 35 items (57%) in the top quarter and on 29 of 35 items in the bottom quarter (83%).

- One item in the top quarter on which the percent correct for all AISD schools exceeded the percent for the nation was "U.S. dropped first atomic bomb on Japan in World War II" (item ranked 16 of 141 nationally).

- On the items in the bottom quarter, all AISD tested exceeded the percent correct for the national on five items:

  --"Union membership grew in the 1930s because of new laws" (item ranked 113 of 141 nationally)

  --"Restrictions on immigration were not part of New Deal" (item ranked 115 of 141 nationally)
  --"Upton Sinclair, Lincoln Steffens, Ida Tarbell known as muckrakers" (item ranked 120 of 141)

  --"Lyndon Johnson's term included Medicare and Voting Rights Act" (item ranked 137 of 141)

  --"'Progressive movement' reforms before World War I" (item ranked 139 of 141)

- On no History item in the top or bottom quarter did the percent correct for the nation exceed the percent correct for every one of the AISD high schools tested.

Literature. In the top quarter, AISD's performance exceeded that of the nation on 23 of 30 items (77%) in the top quarter and on 26 of 30 items in the bottom quarter (87%).

- On the items in the top quarter, all AISD schools tested exceeded the percent correct for the nation on five items:

- --"Martin Luther King, Jr.: 'I have a dream...'" (item ranked 4 of 151 nationally)
- --"Adventures of Huckleberry Finn is about orphan boy and runaway slave" (item ranked 14 of 121)
- --"'Friends, Romans, countrymen...' is from Julius Caesar" (item ranked 20 of 121)
- --"Novel that helped the antislavery movement was Uncle Tom's Cabin" (item ranked 24 of 121)
- --"Lucifer is another name for Satan" (item ranked 27 of 121)

o On the items in the bottom quarter, all AISD schools tested exceeded the percent correct for the nation on six items:

- --"Langston Hughes was poet of Harlem Renaissance who wrote 'Hold fast to dreams,' etc." (item ranked 27 of 121 nationally)
- --"Geoffrey Chaucer wrote The Canterbury Tales" (item ranked 97 of 121)
- --Lord of the Flies is a novel about children stranded on island who try and fail to lead civilized life" (item ranked 100 of 121)
- --"Dante's Divine Comedy is about journey through Hell, Purgatory, Heaven" (item ranked 104 of 120)
- --"Joseph Conrad wrote The Heart of Darkness, Lord Jim, and The Secret Sharer" (item ranked 114 of 120)
- --"Ralph Ellison's Invisible Man is about young man who grows up in South, moves to Harlem" (item ranked 115 of 120)

o On no Literature item in the top or bottom quarter did the percent correct for the nation exceed the percent correct for every one of the AISD high schools tested.

These descriptive analyses suggest that the sample of AISD eleventh graders tested performed better in both history and literature than students in the national sample. Greater percentages of AISD students answered the items answered most frequently by students nationwide, and greater percentages of AISD students answered the items answered least frequently by students in the national sample.

Tests of Significance Between Groups

Using the Statistical Analysis System (SAS) on AISD's IBM mainframe, a one-way analysis of variance (ANOVA) was performed to compare the two groups, AISD and the nation, in terms of the students' mean passing rates on each of the items on the history test. The analysis was repeated for the mean passing rates on each of the items on the literature test. The results are shown in Figures 4 and 5. Figure 6 presents the mean passing rates across items, i.e., the overall average of the percent correct for each item, in history and literature for AISD and the nation.

The results showed significant differences between the groups in both test areas, history and literature.

To determine whether there was a significant difference between paired items, a paired-comparison *t* test was performed. The mean difference between AISD and the nation in history was 6.6014, with a standard error of the mean of 0.7418. This difference was significant (t = 8.90, df = 140, $p < .0001$). In literature, the mean difference between AISD and the nation was 6.3463, with a standard error of the mean of 0.7424. This difference was also significant (t = 8.55, df = 120, $p < .0001$).

Summary

Although problems with representativeness of the AISD sample and with the comparability of the AISD and national samples place limitations on the confidence of the assertion, both descriptive and inferential analyses suggest that the AISD test takers outperformed the 17-year-olds tested nationally. While this is of local interest to AISD administrators, it may also serve as some encouragement to educators in other school districts who when hearing and reading about the national results may have thought as we did, "It can't be *that* bad. I wonder how our students would do on the tests. I'd bet they'd do better." Ours apparently did, although with a 63% passing rate in history and a 56% passing rate in literature (just below and above the failure boundary on Ravitch and Finn's scale), there is clearly room for improvement.

Figure 4. ANOVA SUMMARY TABLE FOR COMPARING THE PERCENT CORRECT SCORES ON THE HISTORY TEST BY AISD AND THE NATION

SOURCE	df	SUM OF SQUARES	MEAN SQUARE	F VALUE
Groups	1	3072.30014184	3072.30014184	9.50*
Error	280	90540.35900710	323.35842503	
Total	281	93612.65914894		

* p < .0023

Figure 5. ANOVA SUMMARY TABLE FOR COMPARING THE PERCENT CORRECT SCORES ON THE LITERATURE TEST BY AISD AND THE NATION

SOURCE	df	SUM OF SQUARES	MEAN SQUARE	F VALUE
Groups	1	2436.65458678	2436.65458678	5.99*
Error	240	97669.05024794	406.95437603	
Total	241	100105.70483472		

* p < .0151

Figure 6. MEAN PASSING RATES ACROSS ALL ITEMS ON THE HISTORY AND LITERATURE TESTS

Test	Group	N*	Mean	Std Dev
History	AISD	141	62.9603	17.7805
History	Nation	141	56.3589	18.1817
Literature	AISD	121	60.4240	19.9321
Literature	Nation	121	54.0777	20.4113

* N = The number of items

Review of the NAEP Item Specifications

As previously stated, the results reported in this paper derive from two sources. This section will focus on the rankings of the NAEP item specifications by local reviewers. The item specifications, representing the domain of content covered by the National Assessment, were chosen for review because much of the controversy about instruction in American history and literature arose in consequence of the national study and the subsequent publication of the results in What Do Our 17-Year-Olds Know?.

As originally envisioned in the operational plan for AISD's study, the review activity was intended to answer the question, "What are the important things which we should teach our children?", the key question on which answers to the other questions depended. Accordingly, it was hoped at first to assemble "persons with a stake in instruction in AISD" to review the item specifications from NAEP before proceeding to the other activities planned. However, activities surrounding the proposed testing of AISD eleventh graders with the NAEP history and literature items assumed such prominence that the reviewers were not selected until late in the 1987-88 school year.

Selection of Reviewers

Two groups of reviewers, one for U.S. history and one for literature, were selected. The reviewers represented teachers, AISD instructional coordinators, and subject area professors from Austin colleges and universities. Forty persons, 20 in each area (U.S. history and literature), were selected to be reviewers. Teacher nominees were solicited within AISD from the secondary instructional coordinators for English and social studies. Nominees among the instructional coordinators were solicited from elementary and secondary administrators. Lists of subject area professors from The University of Texas at Austin, St. Edwards University, Huston-Tillotson College, and Concordia College were drawn up, and selections were made from the lists.

Care was taken in selecting the reviewers to include both men and women and individuals of different ethnicities. Of the 20 reviewers selected for the history item specifications, 11 were men and nine were women; 11 were Anglo, five were Black, and four Hispanic. The 20 reviewers of the literature item specifications were nine men and 11 women, 13 of whom were Anglo, four Black, and three Hispanic.

The Reviewers' Task

Reviewers were sent the NAEP item specifications either for U.S. history or literature and were asked to rate each specification as to its importance for AISD students.

Accompanying the specifications and the rating form was a letter from the Superintendent enlisting their help. Information about the study and the purpose of the review was also sent to each reviewer. The cover letter, general information and the item specifications are Attachments 10, 11, 12, and 13, respectively.

Ten reviewers of the history item specifications and 10 reviewers of the literature specifications responded. Additional efforts were planned in the latter part of summer, 1988, and at the beginning of the 1988-89 school year to persuade the remaining reviewers to submit their ratings. However, budget cuts and a rearrangement of evaluation priorities precluded any follow-up.

Data were tabulated and analyzed using the Statistical Analysis System (SAS) on AISD's IBM mainframe computer. Frequencies of ratings and mean and median ratings were calculated for each objective and subobjective in each curriculum area (see Attachments 12 and 13). Ratings were also calculated by sex and ethnicity.

Results of the Item Specifications Review

Literature. The literature item specifications contained four objectives: Novels, Short Stories and Plays, Myths, Epics, and Biblical Characters and Stories, Poetry, and Nonfiction. (See Attachment 13.) On the 1 to 5 scale, reviewers rated each objective and subobjective between 1 ("essential") and 2 ("important").

Fifty-nine percent of the ratings given were "essential," 31 percent were "important," and five percent were "somewhat important." Only one percent of the ratings were "unimportant" (on the Nonfiction objective). Hispanic and Black reviewers were not as disposed to assign "essential" ratings as were Anglo reviewers; instead, their mean and median ratings favored the "important" and "somewhat important" range. Areas rated "somewhat important" included: Short Stories, Myths, Heroes, and Legends, and Epics. Female reviewers tended to award more favorable ratings than did male reviewers. Their lowest mean and median rating of "important" was for Biblical characters and stories. One of the AISD coordinators felt that Biblical Characters and Stories should be covered by religious institutions. According to this individual, "Teachers are not trained to teach Bible stories in public school settings." Another important issue raised by a reviewer was the need to teach literature to increase the understanding of differences, not just universal characters, plots, and themes. This reviewer felt some emphasis should be given to "the voices of men, women, [and] minorities in addition to the so-called classics."

History. History items were grouped under six objectives: Exploration and Colonization: up to 1763, The Revolutionary War and the New Republic: 1763-1815, Nationhood, Sectionalism, and the Civil War: 1815-1877, The Civil War, Territorial Expansion, the Rise of Modern America, and WWI: 1877-1920, The Great Depression, the New Deal, and WWII: 1920-1945, and Post-WWII: 1945 to Present. Numerous subobjectives were grouped under the objectives. (See Attachment 12.)

Much like their literature reviewer counterparts, all of the mean and median ratings from history reviewers were between 1 and 2, "essential" and "important." One university professor stated, "I find it hard to rate any of these listed objectives less than 'essential'". Objectives and subobjectives receiving 3 ratings ("somewhat important") included:

V. The Great Depression

    D. WWII
    4. The United Nations (e.g., purpose)

VI. Post WWII: 1945-Present

    E. The 1970's
    3. Watergate, resignation of President Nixon, succession of President Ford.

Other areas perceived to be important by the reviewers that they felt were omitted were:

- The life of the first Americans prior to the coming of the Europeans (Professor, UT Austin)

- U.S. involvement in the Middle East and Latin America (Teacher)

- The Hispanic theme in American history (Professor, St. Edwards University).

From a total of 650 responses, 65% and 28%, respectively, were perceived as "essential" and "important". Only 6% of the specifications merited a rating of 3 ("somewhat important"). Females tended to rank more objectives as "essential" (77%) than males (57%). Anglos, who comprised the majority of reviewers, and an Hispanic targeted their ratings in the "essential" or "important" range. The sole Black reviewer was more likely to assign ratings of 2 ("important") and 3 ("somewhat important") than 1 ("essential").

Problems in Carrying Out the Study

Problems with the Test Administration

Although overall the testing went smoothly, there were a couple of problems that inconvenienced school personnel. Some received materials later than school personnel wanted, and some schools received answer sheets sorted differently than expected (despite advance communication from ORE concerning schools' preferences). The worst error was an oversight on the part of ORE staff. The answer sheet which was supplied to schools contained a maximum of five response bubbles. Several items, as it turned out, had six response choices. The difficulty was addressed by providing instructions to schools in marking an unused portion of the answer sheet if the sixth response choice was taken (Attachment 14). The ORE programmer rewrote the scoring programs to accommodate the marking change.

Two other problems are worth noting. The first concerned test security. In materials returned to ORE from one school, one test booklet was not initially accounted for. This missing booklet was extremely problematic in view of the nondisclosure agreement with NAEP. After an exchange of telephone calls and a memo to the principal, the missing booklet was discovered to have been returned improperly. The end result was that all test booklets were accounted for.

The second problem arose from schools not following directions with respect to ensuring that students filled in the test form (see Attachment). Several dozen answer sheets were returned to ORE with the test form left blank. In some instances, ORE staff were able to ascertain which form should have been marked; e.g., L2 had 31 items, while L1, L3, and L4 had 30, and H3 and H4 had 35 and 34 items, respectively, while H1 and H2 had 36 items. However, it was not possible to determine which form had been taken by a student in all cases, and as a result, some answer sheets could not be scored. As mentioned previously, nearly all of one school's answer sheets lacked the form number, so the school did not receive summary results.

These problems notwithstanding, the participating campuses were generally cooperative, communication between ORE and the schools was open and continuous, and the testing appeared to have been well executed.

Resistance to the Study

A study such as this one, which takes for its point of departure a nationally publicized, politically charged issue (how much high school juniors know about U.S. history and world literature, and what to do about it) is bound to encounter difficulties simply because of the emotions roused by the issue. While reactions ran the usual gamut for new studies from "Why are we doing this? This is a stupid idea," to "Testing is fine, but let us choose the students," to "This is an interesting study--I'm curious to find out how it will turn out," this study seemed to encounter more than normal resistance, particularly on the part of local campus personnel.

Although no one challenged the importance of examining the content of the history and literature curricula, there may have been resistance grounded in genuine philosophical differences with the values implicit in the design of the study. Most of the resistance to the study, however, was procedural, i.e., concerned with what was to be done and when. Several rounds of meetings, memos, and phone calls were devoted to questions about which students were to be tested, how much testing time they should be allowed, when and where the testing should take place, what materials would be provided, and who would coordinate the testing. Some of the changes the plans underwent have already been mentioned. Attachments 2 and 3 detail what was finally decided.

Another kind of resistance had at its root a concern that individual school principals and the District as a whole might be embarrassed by the results. Principals evidently saw the testing as a kind of accountability exercise and were wary of the outcomes. The Superintendent was particularly chary of the results potentially becoming another source of negative publicity in the local media, especially in the local newspaper which had adopted something of an adversarial stance with respect to governmental entities in Austin.

Of course, and with some justification, there was objection to any increase in testing from campus administrators and faculty. However, since the testing was originally the Superintendent's idea, the schools' cooperation was expected. In this expectation ORE staff were disappointed, because ultimately, at a principals' meeting at which objections to the testing were voiced, the Superintendent directed that the testing be entirely voluntary. (Curiously, in all of the discussions, other than ORE staff, no one paid much attention to what students might think about the testing and what information they should receive before and after the testing.)

Implications

This study of the content in the curriculum has a number of implications for AISD and for any other school district which might decide to undertake a similar study.

1. Before embarking on a study with strong political overtones, it is imperative that the "front-end" work be completed.

Questions like the following must be thoroughly explored with all of the study's participants:

- What is the purpose of the study?
- Who is requesting that it be conducted and why?
- What resources will be required for the study?
- What will be given up in order to do the study?
- How will the results of the study be used?
- What will the impact on the curriculum be?

These are basic questions, and as experienced research and evaluation staff, we formulated answers which satisfied us, but, in retrospect, the persistent challenges to the study's activities may be seen as indicating that these questions were never entirely resolved for the study's participants on the campuses, despite numerous discussions and written communications.

It is possible that had greater efforts been made to examine these questions in the early going that some resistance to the study would have been dissipated, but it seems to us that the continuing resistance was less a failure of communication on the part of ORE staff than a failure of upper-level administration to support the study beyond its inception.

2. It is necessary that the researchers have the necessary authority and support from the highest administrative levels in the district to carry out a comprehensive study, like this one of curricula or any other.

When the ideas which eventually became the content in the curriculum study first surfaced, there were a number of discussions with the Superintendent and other of the District's top administrators about the study. A comprehensive study design which required the cooperation and assistance of personnel in the instructional departments was cooperatively fashioned, and, as previously mentioned, favorably received by the Superintendent and adopted by the Board of Trustees.

However, after fleshing out the study concept and seeing it underway, the Superintendent and administrators did not, in the end, support the study as originally conceived. Cooperation from administrators in Elementary and Secondary Education was perfunctory, and when high school principals voiced objections to the testing, the Superintendent weighed their objections against the integrity of the design and chose to satisfy the principals.

This was a bitter experience for District researchers. Although it was certainly his prerogative to change his mind, the originator of the study idea, the Superintendent, did not exercise the authority necessary to ensure that a random sample yielding defensibly comparable results would be selected. Instead, in our judgment, the Superintendent was willing to settle for an unknown sample, whose results might be indicative but hardly definitive.

The consequences of this lack of support for the study were several. First, there were opportunity costs. To add the study to the research and evaluation agenda, it was determined that another evaluation would either have to be dropped or reduced. Hence, another evaluation, which was to have been the culmination of a three-year study, was restructured, and certain long-standing questions were left unexamined. Second, the credibility of research and evaluation staff was damaged, and working relationships with campus staff were strained. In short, when administrative support for the study was not forthcoming, the study was drastically impacted and the researchers were in effect "left holding the bag."

In the final analysis, ORE would have done better to have developed a minimum plan to implement the Superintendent's request, rather than have been stuck with an ambitious plan that proved to be beyond the interest and commitment of even the administrators who helped design it.

3. Researchers and administrators are different, and not every study can be completed.

We say this somewhat tongue in cheek, but it is well for researchers occasionally to remember and reflect on the differences between themselves and school administrators. We as researchers customarily view the world as a place where rational decisions can be made based on properly obtained, valid data. Administrators are accustomed to relying on their intuition fed by information from a variety of sources. Researchers are accustomed to explore questions in a methodical, step-by-step fashion through to some resolution. Administrators are readier to take a hopscotch approach, changing the focus of inquiry as circumstances seem to dictate. When an administrator expresses an information need, it is incumbent on the researcher to respond to the request in the context which gave rise to it, and, further, not to be surprised if the need changes.

In AISD, the environment in which the content in the curriculum study was engendered changed. When faced with continuing objections to the testing, the Superintendent decided that the testing did not have to be done as specified in the evaluation design. Administrative priorities in Secondary Education changed and anticipated assistance did not become available. In short, the initial commitment to the study as originally conceived changed. Unfortunately, this change resulted in data of unknown validity and an unfinished study.

If this line of inquiry into what students know or do not know about history and literature (or any other curricular area), and what they should know, (i.e., be taught), is to be productive, future research will need to be more orderly than that attempted in AISD. Judging by the reaction of local reviewers of the NAEP item specifications, there is a great deal of interest among educators in curriculum content. There is likely as keen an interest among the general public. As was intended in AISD's original study design, this is the logical place to start. Testing, while extremely important, should then and only then take place. Obviously, unlike what eventuated in AISD, students should be randomly selected. Finally, there should be a plan for how to present the results, within the district and to the media, and for how the results would be used to improve the teaching and learning situation for the students in the district. Under these conditions, a study such as that attempted in AISD could prove truly useful in curriculum reform, and not merely jumping on a fashionable bandwagon.

ATTACHMENT 1

OATH OF SECURITY AND CONFIDENTIALITY

I, ______________________________, do solemnly swear, or affirm,
(Print full name of principal)
that upon receipt of the assessment materials listed below, I hereby agree that I will not disclose, nor make available to any persons other than those directly connected with the assessment, any part of the assessment materials. These materials are confidential. This oath or affirmation is intended by me to extend to all activities conducted as a part of the administration of the test.

I hereby certify that I have maintained all the security provisions of the assessment. I have neither retained a copy of the test, reproduced a copy of the test or any test item(s), nor divulged to another individual the content of the test or the content or correct answer of any test item.

I certify that I have read this oath or affirmation and that I fully and completely understand the contents. I acknowledge that failure to abide by this, my oath or affirmation, will make me subject to professional sanctions and penalties as can be imposed by law and District policy.

As a testament of this oath I affix my signature below.

Executed this __________ day of _________________________, 19_____.

____
(Signature of Principal)

____
(Campus Name)

ATTACHMENT 2

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

ASSESSMENT OF 11TH GRADERS IN HISTORY AND/OR LITERATURE

Spring, 1988

Building Test Coordinator Checklist

BTC Name:___________________

You should have the following materials from ORE:

- Presorted test booklets (History and/or Literature)
- Answer sheets (preslugged, if requested)
- Fact Sheets for teachers and students (one for each)
- Teacher Checklists (one for you, one for each teacher)
- Oath of Security and Confidentiality (Principal)
- History/Literature Booklets Inventory

You should have the number of answer sheets and booklets you requested plus 10% extra.

Test booklets have been presorted in the order of H1, H2, H3, H4, or L1, L2, L3, and L4, respectively, so that forms will be distributed to students in an approximately equal number. Please note that the "17" preceding the form number has no significance for this testing and should be ignored.

Note that if your school is administering both tests, there will be separate booklets and answer sheets for History and for Literature.

If you do not have the correct amount of materials, or if you need additional materials, please call Letticia Galindo at ORE (458-1227).

NOTE: DO NOT LET STUDENTS DELIVER OR PICK UP TEST MATERIALS!

Prior to Testing

Read the Building Test Coordinator Checklist and the Teacher Checklist.

Record the exact quantities and sequence numbers of test booklets to be given to each teacher on the History/ Literature Booklets Inventory. Do not distribute booklets to teachers until the day of the testing.

Distribute to teachers the Teacher Checklists and Teacher and Student Fact Sheets. Have teachers read these materials and see you if they have any questions.

Convey to the principal the Oath of Security and Confidentiality for the principal's signature. The oath requires the principal to attest that strict test security has been maintained and that no reproduction of test booklets has occurred. It is important that test items remain confidential so that their utility for research purposes is not compromised.

If your school received preslugged answer sheets, teachers will need to check for correctly coded:

- School name and number
- Student name and number
- Teacher name and number

(IF PRESLUGGED INFORMATION IS INCORRECT, DO NOT ATTEMPT TO CHANGE IT. AT THE TOP OF THE ANSWER SHEET, NOTE WHICH INFORMATION IS INCORRECT AND SUPPLY THE CORRECT INFORMATION. ORE WILL CORRECT THE ANSWER SHEET LATER).

If your school did not request preslugged answer sheets, answer sheets will have only school name and number and the type of answer sheet (History or Literature) preslugged. On these answer sheets, students will need to complete student name, student number, and grade.

On all answer sheets, students will need to mark the form of the booklet they receive in row number 172. See Directions for Completing the Student Answer Sheet.

Direct teachers to remove or cover any bulletin board displays or other displays of information that could aid students during testing.

Direct teachers to tell students to bring at least two #2 pencils for the test. Teachers need to maintain extra pencils on hand as well.

Direct teachers to have a watch or clock with a second hand.

Test Day Procedures

Distribute to teachers test booklets and answer sheets. Teachers must initial on the History/Literature Booklets Inventory that they received the booklets specified.

Have teachers pass out test booklets, answer sheets, and Fact Sheets.

Have teachers record starting and ending times of testing on the Teacher Checklist.

The teacher or a proctor should be present in the room during testing.

Suggest to teachers/proctors that they move quietly around the room to make sure students are marking their answers properly on the answer sheet.

Stress that students must not write in their test booklets.

Have teachers tell those students who finish early to check quietly over their work.

After Testing

Have teachers instruct students that after the test, materials are to be arranged in two stacks:

(1) Test booklets
(2) Answer sheets

Instruct teachers that prior to turning materials in to you, they need to do the following:

- Count test booklets
- Check sequence numbers
- Record comments on Teacher Checklist (optional)

Make sure that all materials (including the Oath and Teacher Checklist) are arranged in good order prior to returning to ORE.

PLEASE NOTE: TESTS WILL NOT BE SCORED UNTIL THE OATH OF SECURITY AND CONFIDENTIALITY IS SIGNED BY THE PRINCIPAL AND RETURNED TO ORE.

Mail or deliver materials to ORE.

If you have any questions, please contact Letticia Galindo or David Wilkinson at ORE (458-1227).

HISTORY/LITERATURE BOOKLETS INVENTORY

TO THE BUILDING TEST COORDINATOR

On the day before the test, record the exact quantities and sequence numbers of test booklets given to each teacher. TEACHERS MUST INITIAL THAT THEY HAVE RECEIVED THE MATERIALS SPECIFIED.

TEST BOOKLETS	NUMBER OF BOOKLETS	SEQUENCE NUMBERS	TEACHER	TEACHER INITIALS

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

ASSESSMENT OF 11TH GRADERS IN HISTORY AND/OR LITERATURE

Spring, 1988

Teacher Checklist

Teacher Name:____________________

The Building Test Coordinator should provide you with the following materials:

- * Teacher Checklist, with Directions for Completing the Student Answer Sheet
- * Presorted test booklets (History and/or Literature)
- * Answer sheets (preslugged, if school requested)
- * Fact Sheets for you and students (one for each)

Test booklets have been presorted in the order of H1, H2, H3, H4, or L1, L2, L3, and L4, respectively, so that forms will be distributed to students in an approximately equal number. Please note that the "17" preceding the form number has no significance for this testing and should be ignored.

If you are administering both tests, there will be separate booklets and answer sheets for History and Literature.

NOTE: DO NOT LET STUDENTS DELIVER OR PICK UP TEST MATERIALS!

Prior to Testing

Receive from Building Test Coordinator the Teacher Checklist and Fact Sheets. Read these materials and see the Building Test Coordinator if you have any questions.

If you received preslugged answer sheets, please check for correctly coded:

- o School name and number
- o Student name and number
- o Teacher name and number

(IF PRESLUGGED INFORMATION IS INCORRECT, DO NOT ATTEMPT TO CHANGE IT. AT THE TOP OF THE ANSWER SHEET, NOTE WHICH INFORMATION IS INCORRECT AND SUPPLY THE CORRECT INFORMATION. ORE WILL CORRECT THE ANSWER SHEET LATER.)

If you did not receive preslugged answer sheets, students will need to complete the following prior to the test:

- o Student name and number
- o Grade

On all answer sheets, students will need to mark the form or the booklet they receive in row number 172. See Directions for Completing the Student Answer Sheet.

Remove or cover bulletin board displays or other displays of information that would aid students during testing.

Tell students to bring at least two #2 pencils for the test. Make sure to maintain extra pencils on hand.

Obtain a watch or clock with a second hand.

Test Day Procedures

Receive from Building Test Coordinator test booklets and answer sheets. On the History/Literature Booklets Inventory, you must initial that you have received the booklets specified.

Under no circumstances are you to reproduce any test booklet or disclose the content of any test item. It is important that test items remain confidential so that their utility for research purposes is not compromised.

Pass out test booklets, answer sheets, and Fact Sheets. TEST BOOKLETS ARE ALREADY PRESORTED AND SHOULD BE GIVEN TO STUDENTS IN THIS ORDER: H1, H2, H3, H4, then H1 again, H2, etc. The same sequence should be followed for Literature.

Give students an opportunity to read the Fact Sheet. Please inform them that they can keep these (you can keep yours as well).

Briefly describe for students the testing procedures and after-test procedures. Lead students through the directions for marking test form and checking preslugged information, as apprpriate. See Directions for Completing the Student Answer Sheet.

Stress that students must not write in their test booklets.

Do not assist students inappropriately during the test.

Tell students to erase unwanted answers thoroughly.

If students finish early, instruct them to check quietly over their work.

RECORD STARTING AND ENDING TIMES OF TEST

STARTING	ENDING
_________	_______

Someone should be present in the room during testing.

Move quietly around the room to observe whether students are marking their answers properly on the answer sheet.

After Testing

Once testing is completed, tell students to arrange materials in two stacks: (1) test booklets and (2) answer sheets.

If you are testing in both History and Literature, please collect and stack test booklets and answer sheets from one subject area before passing out booklets and answer sheets for the other test.

Count test booklets: History _______ Literature_______

Check sequence numbers, if necessary:

H1______ H2_______ H3______ H4_______

L1______ L2_______ L3______ L4_______

Please feel free to use the back of this checklist to document any observations you made about the testing process which you think ORE should know.

Turn in all materials (except Fact Sheets) to Building Test Coordinator.

Directions for Completing the Student Answer Sheet

TO THE TEACHER

Please note that if your school requested preslugged answer sheets, they should contain the following information:

* School name and number
* Grade
* Teacher name and number
* Student name and number
* Form (History or Literature)

If your school did not request preslugged answer sheets, they will only contain the following information:

* School name and number
* Form (History or Literature)

A sample answer sheet with the different sections designated has been provided as a guide.

On all answer sheets, students will need to mark the form of the booklet they receive (see below).

Once students receive their answer sheets and test booklets, read the following aloud to the students:

TO THE STUDENT

PLEASE NOTE THE SECTION OF YOUR ANSWER SHEET CONTAINING ROW NUMBERS 171-172. These correspond to the test and booklet you received. If you received a History booklet, this is indicated by an "A" in 171; a "B" if it is a Literature booklet. There are a total of 8 booklets (4 History and 4 Literature) which are designated H1, H2, H3, H4 and L1, L2, L3, and L4, respectively.

In row number 172:

Mark "A" if you received Booklet H1 or L1.
Mark "B" if you received Booklet H2 or L2.
Mark "C" if you received Booklet H3 or L3.
Mark "D" if you received Booklet H4 or L4.

It is essential for scoring purposes that the wrong test key not be mistakenly used because booklets were not marked properly. In addition, this information serves to help ORE keep track of booklets and answer sheets for all students tested.

IF YOU ARE TAKING THE **HISTORY** ASSESSMENT, BEGIN MARKING YOUR ANSWER SHEET WITH # 13.

IF YOU ARE TAKING THE **LITERATURE** ASSESSMENT, BEGIN MARKING YOUR ANSWER SHEET WITH # 19.

Directions for Completing the Student Answer Sheet

TO THE TEACHER

STUDENTS ARE TO SKIP THE FOLLOWING QUESTIONS ACCORDING TO BOOKLET AND FORM NUMBER:

H1--------------- 1-12, 49-61
H2--------------- 1-12, 49-61
H3--------------- 1-12, 48-60
H4--------------- 1-12, 47-59

L1--------------- 1-18, 49-72
L2--------------- 1-18, 50-73
L3--------------- 1-18, 49-72
L4--------------- 1-18, 49-72

Please check to see that students have marked the correct form numbers on their answer sheets and that they have marked their answers in the proper spaces.

ATTACHMENT 4

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

ASSESSMENT OF 11TH GRADERS IN U.S. HISTORY AND/OR LITERATURE

Spring, 1988

Fact Sheet

TO THE TEACHER:

About the Test

You are about to administer a test either of U.S. history or of literature. There are four forms of each test divided up among four booklets. The history test booklets are labeled H1, H2, H3, and H4. The literature test booklets are labeled L1, L2, L3, and L4. Booklets should be passed out to students in this order. Depending on which form the student receives, each student will answer from 34-36 history questions and/or 30-31 literature questions.

Preceding and following the knowledge component of the test are approximately 11-12 questions pertaining to coursework. The students should skip these items.

Estimated working time for each test (history and literature) is approximately 14 minutes but a total of 20 minutes should be allotted to ensure students will have ample time to take the test (40 minutes if both tests are given).

Why It's Being Given

The purpose of the test is to determine whether AISD students are learning important content in American history and literature. Although the test will not be used to determine student's grades in these subjects or affect their graduation status, the results should provide your school with valuable information about what is being taught and what needs to be taught in these subject areas. Therefore, please encourage the students to give it their best effort.

Testing materials and directions should be obtained from the Building Test Coordinator.

Thank you for your cooperation in this important testing effort!

ATTACHMENT 5

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

ASSESSMENT OF 11TH GRADERS IN U.S. HISTORY AND/OR LITERATURE

Spring, 1988

Fact Sheet

TO THE STUDENT:

The test you are about to take consists of questions designed to measure your knowledge of U.S. history and/or literature. At the beginning and end of the test, there are approximately 11-12 questions that deal with coursework in these subject areas. Skip these questions.

The estimated working time for the test is approximately 14 minutes, but you will be allowed 20 minutes (40 minutes if both tests are given).

The purpose of the test is to determine whether AISD students are learning important content in American history and literature and to compare their performance with the performance of students nationwide. Although the test will not be used to determine your grade in these subjects or affect your graduation status, the results should provide your school with valuable information about what is being taught and what needs to be taught in these subject areas. Therefore, please give it your best effort.

THANK YOU FOR YOUR COOPERATION. GOOD LUCK!

ATTACHMENT 6
STANDARD ANSWER SHEET

SCHOOL NAME CROCKETT

SCHOOL NUMBER 008

INSTRUCTOR BOWER

MARKING INSTRUCTIONS

USE NO. 2 PENCIL ONLY

CORRECT MARKING

A CORRECT MARK SHOULD COVER THE COMPLETE OUTLINE

EXAMPLE	
○ ○ ○ ● ○ ○	2
○ ● ○ ○ ○ ○	4
○ ○ ● ○ ○ ○	3
● ○ ○ ○ ○ ○	5
○ ○ ○ ○ ○ ●	0

○ ○ ○ ○ ○ ○ ○ ○ ○ ●	0	L
○ ○ ○ ○ ○ ○ ○ ○ ○ ●	0	O
○ ● ○ ○ ○ ○ ○ ○ ○ ○	8	C
○ ○ ○ ○ ○ ○ ○ ● ○ ○	2	T
○ ○ ○ ○ ● ○ ○ ○ ○ ○	5	C
○ ○ ○ ○ ○ ○ ● ○ ○ ○	3	H
○ ○ ○ ● ○ ○ ○ ○ ○ ○ ○ ○		GRADE

[redacted]

[redacted]

A
D
R
I
A
N
J

1	☒ ☒ ☒ ☒ ☒										
2	☒ ☒ ☒ ☒ ☒										
3	☒ ☒ ☒ ☒ ☒										
4	☒ ☒ ☒ ☒ ☒										
5	☒ ☒ ☒ ☒ ☒										
6	☒ ☒ ☒ ☒ ☒										
7	☒ ☒ ☒ ☒ ☒										
8	☒ ☒ ☒ ☒ ☒										
9	☒ ☒ ☒ ☒ ☒										
10	☒ ☒ ☒ ☒ ☒										
11	☒ ☒ ☒ ☒ ☒										
12	☒ ☒ ☒ ☒ ☒										
13	○ ○ ○ ● ○										
14	● ○ ○ ○ ○										
15	○ ○ ● ○ ○										
16	○ ○ ● ○ ○										
17	● ○ ○ ○ ○										
18	○ ● ○ ○ ○										
19	● ○ ○ ○ ○										
20	○ ● ○ ○ ○										
21	● ○ ○ ○ ○	51	○ ○ ○ ○ ○	81	○ ○ ○ ○ ○	111	○ ○ ○ ○ ○	141	○ ○ ○ ○ ○	171	● ○ ○ ○ ○
22	○ ○ ● ○ ○	52	○ ○ ○ ○ ○	82	○ ○ ○ ○ ○	112	○ ○ ○ ○ ○	142	○ ○ ○ ○ ○	172	○ ○ ○ ○ ○
23	● ○ ○ ○ ○	53	○ ○ ○ ○ ○	83	○ ○ ○ ○ ○	113	○ ○ ○ ○ ○	143	○ ○ ○ ○ ○	173	○ ○ ○ ○ ○
24	○ ○ ○ ● ○	54	○ ○ ○ ○ ○	84	○ ○ ○ ○ ○	114	○ ○ ○ ○ ○	144	○ ○ ○ ○ ○	174	○ ○ ○ ○ ○
25	○ ○ ○ ● ○	55	○ ○ ○ ○ ○	85	○ ○ ○ ○ ○	115	○ ○ ○ ○ ○	145	○ ○ ○ ○ ○	175	○ ○ ○ ○ ○
26	○ ● ○ ○ ○	56	○ ○ ○ ○ ○	86	○ ○ ○ ○ ○	116	○ ○ ○ ○ ○	146	○ ○ ○ ○ ○	176	○ ○ ○ ○ ○
27	○ ○ ● ○ ○	57	○ ○ ○ ○ ○	87	○ ○ ○ ○ ○	117	○ ○ ○ ○ ○	147	○ ○ ○ ○ ○	177	○ ○ ○ ○ ○
28	● ○ ○ ○ ○	58	○ ○ ○ ○ ○	88	○ ○ ○ ○ ○	118	○ ○ ○ ○ ○	148	○ ○ ○ ○ ○	178	○ ○ ○ ○ ○
29	○ ○ ● ○ ○	59	○ ○ ○ ○ ○	89	○ ○ ○ ○ ○	119	○ ○ ○ ○ ○	149	○ ○ ○ ○ ○	179	○ ○ ○ ○ ○
30	○ ● ○ ○ ○	60	○ ○ ○ ○ ○	90	○ ○ ○ ○ ○	120	○ ○ ○ ○ ○	150	○ ○ ○ ○ ○	180	○ ○ ○ ○ ○
31	○ ○ ○ ● ○	61	○ ○ ○ ○ ○	91	○ ○ ○ ○ ○	121	○ ○ ○ ○ ○	151	○ ○ ○ ○ ○	181	○ ○ ○ ○ ○
32	● ○ ○ ○ ○	62	○ ○ ○ ○ ○	92	○ ○ ○ ○ ○	122	○ ○ ○ ○ ○	152	○ ○ ○ ○ ○	182	○ ○ ○ ○ ○
33	○ ○ ● ○ ○	63	○ ○ ○ ○ ○	93	○ ○ ○ ○ ○	123	○ ○ ○ ○ ○	153	○ ○ ○ ○ ○	183	○ ○ ○ ○ ○
34	○ ○ ○ ● ○	64	○ ○ ○ ○ ○	94	○ ○ ○ ○ ○	124	○ ○ ○ ○ ○	154	○ ○ ○ ○ ○	184	○ ○ ○ ○ ○
35	○ ○ ○ ● ○	65	○ ○ ○ ○ ○	95	○ ○ ○ ○ ○	125	○ ○ ○ ○ ○	155	○ ○ ○ ○ ○	185	○ ○ ○ ○ ○
36	○ ● ○ ○ ○	66	○ ○ ○ ○ ○	96	○ ○ ○ ○ ○	126	○ ○ ○ ○ ○	156	○ ○ ○ ○ ○	186	○ ○ ○ ○ ○
37	○ ○ ● ○ ○	67	○ ○ ○ ○ ○	97	○ ○ ○ ○ ○	127	○ ○ ○ ○ ○	157	○ ○ ○ ○ ○	187	○ ○ ○ ○ ○
38	○ ○ ● ○ ○	68	○ ○ ○ ○ ○	98	○ ○ ○ ○ ○	128	○ ○ ○ ○ ○	158	○ ○ ○ ○ ○	188	○ ○ ○ ○ ○
39	○ ● ○ ○ ○	69	○ ○ ○ ○ ○	99	○ ○ ○ ○ ○	129	○ ○ ○ ○ ○	159	○ ○ ○ ○ ○	189	○ ○ ○ ○ ○
40	○ ○ ● ○ ○	70	○ ○ ○ ○ ○	100	○ ○ ○ ○ ○	130	○ ○ ○ ○ ○	160	○ ○ ○ ○ ○	190	○ ○ ○ ○ ○
41	○ ● ○ ○ ○	71	○ ○ ○ ○ ○	101	○ ○ ○ ○ ○	131	○ ○ ○ ○ ○	161	○ ○ ○ ○ ○	191	○ ○ ○ ○ ○
42	● ○ ○ ○ ○	72	○ ○ ○ ○ ○	102	○ ○ ○ ○ ○	132	○ ○ ○ ○ ○	162	○ ○ ○ ○ ○	192	○ ○ ○ ○ ○
43	○ ○ ○ ● ○	73	○ ○ ○ ○ ○	103	○ ○ ○ ○ ○	133	○ ○ ○ ○ ○	163	○ ○ ○ ○ ○	193	○ ○ ○ ○ ○
44	○ ○ ● ○ ○	74	○ ○ ○ ○ ○	104	○ ○ ○ ○ ○	134	○ ○ ○ ○ ○	164	○ ○ ○ ○ ○	194	○ ○ ○ ○ ○
45	○ ● ○ ○ ○	75	○ ○ ○ ○ ○	105	○ ○ ○ ○ ○	135	○ ○ ○ ○ ○	165	○ ○ ○ ○ ○	195	○ ○ ○ ○ ○
46	○ ● ○ ○ ○	76	○ ○ ○ ○ ○	106	○ ○ ○ ○ ○	136	○ ○ ○ ○ ○	166	○ ○ ○ ○ ○	196	○ ○ ○ ○ ○
47	○ ○ ○ ● ○	77	○ ○ ○ ○ ○	107	○ ○ ○ ○ ○	137	○ ○ ○ ○ ○	167	○ ○ ○ ○ ○	197	○ ○ ○ ○ ○
48	○ ○ ○ ● ○	78	○ ○ ○ ○ ○	108	○ ○ ○ ○ ○	138	○ ○ ○ ○ ○	168	○ ○ ○ ○ ○	198	○ ○ ○ ○ ○
49	○ ○ ○ ○ ○	79	○ ○ ○ ○ ○	109	○ ○ ○ ○ ○	139	○ ○ ○ ○ ○	169	○ ○ ○ ○ ○	199	○ ○ ○ ○ ○
50	○ ○ ○ ○ ○	80	○ ○ ○ ○ ○	110	○ ○ ○ ○ ○	140	○ ○ ○ ○ ○	170	○ ○ ○ ○ ○	200	○ ○ ○ ○ ○

TY
FO

H
I
S
T
O
R
Y

EB14961-54321 A1904-4

ATTACHMENT 7

STANDARD ANSWER SHEET

S071

SCHOOL NAME JOHNSTON

SCHOOL NUMBER 003

INSTRUCTOR MIZE

MARKING INSTRUCTIONS

USE NO. 2 PENCIL ONLY

CORRECT MARKING

EXAMPLE

2

4

3

5

0

A CORRECT MARK SHOULD COVER THE COMPLETE OUTLINE

0 L

0 O

3 C

6 T

2 C

0 H

GRADE

[redacted]

[redacted]

M

A

R

T

A

I

TYP FOR

L
I
T
E
R
A
T
U
R
E

```
     PROGRAM: SA$NAEPB                            .    AUSTIN INDEPEN
                                                     DEPARTMENT OF M
                                                      OFFICE OF RESE

                                             DISTRICTWIDE HISTORY & LIT

------------------------------------------------------------------ TES
National   AISD
  Rank     Rank     ITEM
                    DESCRIPTION

     1        4     THOMAS EDISON INVENTED THE LIGHT BULB
     2        1     FIND SOVIET UNION ON MAP OF EUROPE
     3        8     ALEXANDER GRAHAM BELL INVENTED THE TELEPHO
     4       27     GEORGE WASHINGTON WAS PRESIDENT BETWEEN 17
     5        2     FIND ITALY ON MAP OF EUROPE
     6        6     THE UNDERGROUND RAILROAD WAS SECRET NETWOR
     7       12     ADOLF HITLER WAS LEADER OF GERMANY DURING
     8       21     THOMAS JEFFERSON WAS MAIN AUTHOR OF DECLAR
   . 9       14     THE ASSEMBLY LINE WAS INTRODUCED BY US AUT
    10       26     FIND AREA OF 13 ORIGINAL STATES ON MAP OF
    11        3     KKK USED VIOLENCE TO OPPOSE MINORITIES
    12       13     HARRIET TUBMAN WAS LEADER IN HELPING SLAVE
    13       37     FIND ROCKY MOUNTAINS ON MAP OF US
    14       19     GUARANTEE OF FREE SPEECH AND RELIGION FOUN
    15       15     JAPANESE ATTACK ON PEARL HARBOR LED US INT
    16        7     US DROPPED FIRST ATOMIC BOMB ON JAPAN IN W
    17       33     WASHINGTON WAS COMMANDER OF AMERICAN ARMY
    18       23     CHURCHILL WAS PRIME MINISTER OF GREAT BRIT
    19       41     TV BECAME NEW FEATURE IN AMERICAN HOMES AF
    20       40     WATERGATE LED TO RESIGNATION OF RICHARD NI
    21       24     WOMEN WORKED IN FACTORIES DURING WW II
    22       50     JAMESTOWN WAS FIRST PERMANENT ENGLISH COLO
    23       20     FIND WEST GERMANY ON MAP OF EUROPE
    24       42     LINDBERGH MADE FIRST SOLO FLIGHT ACROSS AT
    25       36     NAZI DECIMATION OF JEWISH PEOPLE IN EUROPE
    26       16     THE GREAT DEPRESSION WAS A PERIOD OF MASS
    27       10     "PROHIBITION" REFERS TO BAN ON SALE AND CO
    28       34     THE GREAT DEPRESSION OCCURRED BETWEEN 1900
    29       47     VIETNAM WAR HAD STRONGEST CITIZEN PROTEST
    30       25     CIVIL RIGHTS MOVEMENT OF 1960'S FOCUSED ON
    31       53     DISCOVERY OF GOLD IN CALIFORNIA WAS A CAUS
    32        9     FIND AREA ON US MAP THAT FOUGHT FOR INDEPE
    33       35     INVENTION OF COTTON GIN STIMULATED PLANTAT
    34       22     GERMANY AND JAPAN WERE MAIN US ENEMIES IN
    35       31     WW II ENDED BETWEEN 1943-1947
    36        5     INDIANS WERE PUT ON RESERVATIONS AFTER CIV
    37       72     FIND MISSISSIPPI RIVER ON MAP OF US
    38       11     FIND GREAT BRITAIN ON MAP OF EUROPE
    39       69     MOST PEOPLE IN ENGLISH COLONIES OF NORTH A
    40       17     "SECESSION" REFERS TO WITHDRAWAL OF SOUTHE
    41       56     SUSAN B. ANTHONY WAS A LEADER OF WOMEN'S S
    42       64     COLUMBUS DISCOVERED THE NEW WORLD BEFORE 1
    43       32     ABRAHAM LINCOLN WROTE THE EMANCIPATION PRO
    44       65     THE DECLARATION OF INDEPENDENCE WAS SIGNED
    45       51     THE DECLARATION OF INDEPENDENCE MARKS COLO
    46       46     STAMP ACT OF 1765 WAS ATTEMPT BY ENGLISH T
    47       49     US POLICY AFTER WW II WAS TO PROVIDE AID T
    48       30     FIND FRANCE ON MAP OF EUROPE
    49       18     US PROVOKED TO ENTER WW I BY GERMAN SUBMAR
```

4

DENT SCHOOL DISTRICT 8:01 FRIDAY, AUGUST 19, 1988 1
ANAGEMENT INFORMATION
ARCH AND EVALUATION

ERATURE TEST RANK ORDER RESULTS

T=HISTORY --

	# STUDENTS RECEIVING ITEM	# CORRECT	% CORRECT
	173	154	89.0
	173	167	96.5
NE	176	154	87.5
30-1800	173	140	80.9
	173	160	92.5
K TO HELP SLAVES ESCAPE	176	156	88.6
WW II	172	150	87.2
ATION OF INDEPENDENCE	174	143	82.2
D INDUSTRY	174	149	85.6
US	174	141	81.0
	173	156	90.2
S ESCAPE	172	149	86.6
	172	133	77.3
D IN BILL OF RIGHTS	174	144	82.8
D WW II	173	147	85.0
W II	174	153	87.9
IN REVOLUTION	172	135	78.5
AIN IN WW II	173	141	81.5
TER 1950	174	132	75.9
XON	172	131	76.2
	173	141	81.5
NY	173	124	71.7
	173	143	82.7
LANTIC	174	132	75.9
CALLED THE HOLOCAUST	174	135	77.6
UNEMPLOYMENT	173	146	84.4
NSUMPTION OF LIQUOR	173	151	87.3
-1950	172	134	77.9
MOVEMENTS	174	128	73.6
EQUALITY FOR MINORITIES	176	143	81.3
E OF WESTWARD MOVEMENT	172	122	70.9
NDENCE FROM MEXICO	176	154	87.5
ION ECONOMY IN SOUTH	172	134	77.9
WW II	176	144	81.8
	176	139	79.0
IL WAR	173	154	89.0
	172	113	65.7
	173	151	87.3
MERICA WORKED AS FARMERS	173	115	66.5
RN STATES FROM UNION	174	145	83.3
UFFRAGE MOVEMENT	174	122	70.1
750	174	117	67.2
CLAMATION	176	139	79.0
BETWEEN 1750-1800	172	115	66.9
NIES' BREAK W/ ENGLAND	176	126	71.6
O TAX COLONIES	172	127	73.8
N EUROPE	173	125	72.3
	173	137	79.2
INE ATTACKS	172	143	83.1

ATTACHMENT 8

```
PROGRAM: SA$NAEP8                             AUSTIN INDEPEND
                                           DEPARTMENT OF MA
                                             OFFICE OF RESEA

                                   DISTRICTWIDE HISTORY & LITE

-------------------------------------------------------- TEST
National   AISD
  Rank     Rank     ITEM
                  DESCRIPTION

   50      108    WATERGATE OCCURRED AFTER 1950
   51       45    BROWN DECISION OF 1954 RULED SCHOOL SEGREGA
   52       74    ALEXANDER HAMILTON KNOWN FOR URGING STRONG
   53       63    FIND AREA OF THE CONFEDERACY ON MAP OF US
   54       38    SPUTNIK WAS THE FIRST MAN-MADE SATELLITE
   55       48    PLESSY VS. FERGUSON DECISION APPROVED RACIA
   56       88    THE SOUTHWESTERN US WAS EXPLORED AND SETTLE
   57       60    UNITED STATES CONSTITUTION WAS WRITTEN BETW
   58       70    EAST COAST OF US EXPLORED AND SETTLED MAINL
   59       28    WOODROW WILSON APPEALED FOR AMERICAN ENTRY
   60       54    JAPAN BOMBED PEARL HARBOR BETWEEN 1939-1943
   61       71    "CHECKS AND BALANCES" DIVIDES POWER AMONG B
   62       77    HEAVY IMMIGRATION IN 1970S-80S CAME FROM SE
   63       85    ARTICLES OF CONFEDERATION WERE 1ST CONSTITU
   64       57    REPRESENTATIVE GOVERNMENT EXISTED IN COLONI
   65       78    JOHN D. ROCKEFELLER FORMED STANDARD OIL
   66       66    US FOREIGN POLICY AFTER WW II WAS CONTAINME
   67       73    WW II WAS BETWEEN 1900-1950
   68      121    BOOKER T. WASHINGTON WAS MAJOR BLACK LEADER
   69       52    FIND AREA OF MAP OF US PURCHASED FROM FRANC
   70       75    "PROHIBITION" OCCURRED BETWEEN 1900-1950
   71       84    BEFORE 1800 MOST IMMIGRANTS CAME FROM NORTH
   72       55    EISENHOWER WAS PRESIDENT BETWEEN 1946-1963
   73       81    ISRAEL HAS NEVER BEEN INVADED BY SOVIET UNI
   74       62    JAPANESE-AMERICANS WERE FORCED INTO CAMPS D
   75      102    NIXON ESTABLISHED RELATIONS WITH THE PEOPLE
   76       79    A CAUSE OF POPULATION MOVEMENT IN 1800S WAS
   77       43    SOCIAL SECURITY SYSTEM WAS INTRODUCED DURIN
   78       68    US GAVE MILITARY SUPPORT TO SOUTH KOREA WIT
   79       29    DUST BOWL OF 1930S FOSTERED MOVEMENT TO CAL
   80       76    AMENDMENT OF 1920 GRANTED WOMEN RIGHT TO VO
   81       61    STALIN WAS LEADER OF SOVIET UNION DURING WW
   82       97    POPULIST PARTY ADVOCATED GOVERNMENT SUPPORT
   83       58    NEW DEAL IDENTIFIED WITH CHANGES IN SOCIAL,
   84       44    FRANKLIN ROOSEVELT WAS PRESIDENT BETWEEN 19
   85       86    BEFORE CIVIL WAR, NATION DEBATED SPREAD OF
   86      124    "GIVE ME LIBERTY OR GIVE ME DEATH"--PATRICK
   87       39    "LAISSEZ-FAIRE" MEANS MINIMAL GOVERNMENT RE
   88       95    MONROE DOCTRINE: EUROPE SHOULDN'T GAIN LAND
   89      101    WASHINGTON'S FAREWELL ADDRESS WARNED AGAINS
   90       87    CANADA AND MISSISSIPPI VALLEY FIRST EXPLORE
   91      114    SAMUEL GOMPERS WAS FIRST PRESIDENT OF AFL
   92       82    MLK, JR., ROSE TO PROMINENCE IN MONTGOMERY
   93       91    US DROPPED FIRST ATOMIC BOMB BETWEEN 1943-1
   94       96    VALLEY FORGE WAS LOWEST POINT IN REVOLUTION
   95      105    WOMEN'S RIGHT TO VOTE WAS NOT GUARANTEED AF
   96      106    AN ISSUE IN WAR OF 1812: GREAT BRITAIN'S IN
   97      103    ANDREW CARNEGIE WAS ASSOCIATED WITH DEVELOP
   98      113    THOMAS JEFFERSON WAS PRESIDENT BETWEEN 1800
```

ENT SCHOOL DISTRICT 8:01 FRIDAY, AUGUST 19, 1988 2
NAGEMENT INFORMATION
RCH AND EVALUATION

RATURE TEST RANK ORDER RESULTS

=HISTORY --

	# STUDENTS RECEIVING ITEM	# CORRECT	% CORRECT
	174	82	47.1
TION UNCONSTITUTIONAL	174	129	74.1
NATIONAL GOVERNMENT	173	113	65.3
	173	118	68.2
	172	133	77.3
L SEGREGATION	172	125	72.7
D BY SPAIN	174	104	59.8
EEN 1750-1800	174	120	69.0
Y BY ENGLAND	176	116	65.9
INTO LEAGUE OF NATIONS	173	139	80.3
	176	124	70.5
RANCHES OF FED'L GOV'T.	173	114	65.9
ASIA AND LATIN AMERICA	173	107	61.8
TION DURING REVOLUTION	174	105	60.3
ES	176	123	69.9
	172	106	61.6
NT OF COMMUNISM	172	115	66.9
	174	114	65.5
BEFORE WW I	174	70	40.2
E	176	125	71.0
	174	113	64.9
ERN AND WESTERN EUROPE	172	104	60.5
	172	121	70.3
ON	172	105	61.0
URING WW II	172	118	68.6
'S REPUBLIC OF CHINA	173	88	50.9
REASONABLE LAND PRICES	174	107	61.5
G THE NEW DEAL	176	132	75.0
H UNITED NATIONS	176	117	66.5
IFORNIA	176	141	80.1
TE	172	111	64.5
II	174	120	69.0
FOR FARMERS	176	95	54.0
ECONOMIC POLICIES	174	121	69.5
29-1946	172	129	75.0
SLAVERY	173	104	60.1
HENRY	176	70	39.8
GULATION OF ECONOMY	174	133	76.4
IN WESTERN HEMISPHERE	173	96	55.5
T FOREIGN ALLIANCES	176	90	51.1
D BY FRANCE	172	103	59.9
	176	79	44.9
BUS BOYCOTT	172	105	61.0
947	176	104	59.1
ARY WAR	174	94	54.0
TER CIVIL WAR	173	83	48.0
TERFERENCE W/ SHIPPING	172	82	47.7
MENT OF STEEL INDUSTRY	176	88	50.0
-1820	173	78	45.1

PROGRAM: SA$NAEP8

AUSTIN INDEP
DEPARTMENT OF
OFFICE OF RE

DISTRICTWIDE HISTORY & L

- T

National Rank	AISD Rank	ITEM DESCRIPTION
99	99	THE CONSTITUTION DIVIDES POWERS BETWEEN
100	89	MISSOURI COMPROMISE ADMITTED MAINE AS FR
101	83	WOODROW WILSON WAS PRESIDENT BETWEEN 191
102	59	SENATOR JOSEPH MCCARTHY INVOLVED IN CONT
103	118	"NULLIFICATION" RELATED TO STATES' RIGHT
104	92	HOOVER, FRANKLIN ROOSEVELT WERE PRESIDEN
105	128	JANE ADDAMS FOUNDED SETTLEMENT HOUSES TO
106	109	"RECONSTRUCTION" OCCURRED BETWEEN 1850-1
107	126	"THE FEDERALIST" ADVOCATED ADOPTION OF C
108	125	DRED SCOTT DECISION: SLAVE WHO MOVED TO
109	137	D-DAY OCCURRED BETWEEN 1943-1947
110	98	RENAISSANCE WAS CHARACTERIZED BY CULTURA
111	135	PAINE'S "COMMON SENSE" ARGUES FOR COLONI
112	112	EMANCIPATION PROCLAMATION FREED SLAVES I
113	93	UNION MEMBERSHIP GREW IN THE 1930S BECAU
114	122	JAMESTOWN FOUNDED BEFORE 1750
115	67	RESTRICTIONS ON IMMIGRATION WERE NOT PAR
116	119	"3/5 COMPROMISE" IN CONSTITUTION DEFINED
117	111	IMMIGRATION FROM SOUTHERN AND EASTERN EU
118	110	IMMIGRATION RESTRICTION IN 1921, 1924 AI
119	100	SCOPES TRIAL WAS ABOUT TEACHING EVOLUTIO
120	90	UPTON SINCLAIR, LINCOLN STEFFENS, IDA TA
121	107	THEODORE ROOSEVELT WAS PRESIDENT BETWEEN
122	133	ARTICLES OF CONFEDERATION FAILED TO PROV
123	132	FIND REGION ON US MAP ACQUIRED FROM MEXI
124	104	RELIGIOUS TOLERATION IN COLONIES DUE TO
125	120	JONAS SALK INVENTED POLIO VACCINE
126	117	SPANISH-AMERICAN WAR MADE US AN INTERNAT
127	80	AMERICAN FOREIGN POLICY AFTER WW I KNOWN
128	123	THE CIVIL WAR OCCURRED BETWEEN 1850-1900
129	115	US FOREIGN POLICY IN EARLY 1900S:"SPEAK
130	127	PURPOSE OF JIM CROW LAWS WAS TO ENFORCE
131	134	MAGNA CARTA IS FOUNDATION OF BRITISH PAR
132	129	ANDREW JACKSON WAS PRESIDENT BETWEEN 182
133	130	REFORMATION LED TO ESTABLISHMENT OF PROT
134	116	THE UNITED NATIONS WAS FOUNDED BETWEEN 1
135	136	THE SENECA FALLS DECLARATION CONCERNED W
136	141	ABRAHAM LINCOLN WAS PRESIDENT BETWEEN 18
137	94	LYNDON JOHNSON'S TERM INCLUDED MEDICARE
138	138	BETTY FRIEDAN AND GLORIA STEINEM LED WOM
139	131	"PROGRESSIVE MOVEMENT" REFERS TO REFORMS
140	140	"RECONSTRUCTION" REFERS TO READMISSION O
141	139	JOHN WINTHROP AND THE PURITANS FOUNDED A

```
ENDENT SCHOOL DISTRICT                        8:01 FRIDAY, AUGUST 19, 1988    3
 MANAGEMENT INFORMATION
SEARCH AND EVALUATION

ITERATURE TEST RANK ORDER RESULTS

EST=HISTORY -----------------------------------------------------------------

                                  # STUDENTS            #            %
                                RECEIVING ITEM       CORRECT      CORRECT

STATES AND FED'L. GOV'T.             172                90          52.3
EE, MO. AS SLAVE STATE               174               104          59.8
2-1929                               172               105          61.0
ROVERSY ABOUT COMMUNISM              174               121          69.5
S                                    173                72          41.6
TS DURING DEPRESSION                 176               104          59.1
 HELP THE POOR                       176                67          38.1
900                                  174                82          47.1
ONSTITUTION                          176                69          39.2
FREE STATE WAS NOT FREE              176                70          39.8
                                     176                54          30.7
., TECHNOLOGICAL ADVANCES            174                92          52.9
AL INDEPENDENCE                      173                57          32.9
N CONFEDERACY                        172                79          45.9
SE OF NEW LAWS                       174               100          57.5
                                     174                70          40.2
T OF NEW DEAL                        172               115          66.9
 STATUS OF SLAVES                    173                72          41.6
ROPE GREW, 1890-1910                 176                81          46.0
MED AT SE EUROPEANS                  174                82          47.1
N                                    172                90          52.3
RBEL KNOWN AS MUCKRAKERS             172               102          59.3
 1895-1912                           172                82          47.7
IDE ADEQUATE TAXING POWER            173                58          33.5
CO IN WAR                            176                62          35.2
COMMON INTEREST OF MANY              176                87          49.4
                                     176                71          40.3
IONAL POWER                          172                74          43.0
 AS ISOLATIONIST                     174               107          61.5
                                     172                69          40.1
SOFTLY, CARRY A BIG STICK"           176                76          43.2
RACIAL SEGREGATION                   174                68          39.1
LIAMENTARY SYSTEM                    173                58          33.5
)-1840                               173                64          37.0
ESTANT SECTS                         172                62          36.0
943-1947                             176                76          43.2
OMEN'S RIGHTS                        176                55          31.3
50-1880                              173                35          20.2
AND VOTING RIGHTS ACT                173                98          56.6
EN'S MOVEMENT IN 1970S               176                44          25.0
 BEFORE WW I                         173                62          35.8
 CONFEDERATE STATES                  176                42          23.9
 COLONY IN BOSTON                    172                43          25.0
```

PROGRAM: SA$NAEP8

AUSTIN INDEPEN
DEPARTMENT OF M
OFFICE OF RESE

DISTRICTWIDE HISTORY & LIT

-- TEST=

National Rank	AISD Rank	ITEM DESCRIPTION
1	3	IN BIBLE, NOAH GATHERED PAIRS OF CREATURES
2	5	IN BIBLE, MOSES GOT 10 COMMANDMENTS, LED H
3	4	ROMEO AND JULIET HINDERED BY FEUDING FAMIL
4	1	MLK, JR.: "I HAVE A DREAM..."
5	10	HAMLET: "TO BE OR NOT TO BE..." (5 LINES)
6	8	EBENEZER SCROOGE WAS STINGY CHARACTER IN "
7	6	IN GREEK MYTHOLOGY, RULER OF GODS IS ZEUS
8	12	WHITE RABBIT, MARCH HARE, ETC., ARE IN ALI
9	7	ROBIN HOOD IS KNOWN FOR STEALING FROM RICH
10	13	CINDERELLA'S RAGS TURNED INTO GOWN AND SHE
11	17	ROBINSON CRUSOE WAS SHIPWRECKED ON AN ISLA
12	14	"THE LORD IS MY SHEPHERD..." (3 LINES) IS
13	18	THE MAGICIAN WHO ADVISED KING ARTHUR WAS M
14	2	ADVENTURES OF HUCKLEBERRY FINN: ORPHAN BOY
15	21	IN BIBLE, BOOK OF GENESIS INCLUDES ACCOUNT
16	15	PLATO AND ARISTOTLE ARE BEST KNOWN AS PHIL
17	23	VICTORY OF UNKNOWN TEAM PARALLELS BIBLE ST
18	25	"RIP VAN WINKLE" IS ABOUT MAN WHO SLEPT FO
19	29	E.A. POE WROTE "PIT AND PENDULUM," "FALL O
20	9	"FRIENDS, ROMANS, COUNTRYMEN..." IS FROM "
21	26	JULIET SAID, "WHAT'S IN A NAME? THAT WHIC
22	33	GETTYSBURG ADDRESS: "FOUR SCORE & SEVEN YE
23	27	SHELLEY'S FRANKENSTEIN: SCIENTIST MADE CRE
24	22	NOVEL THAT HELPED ANTISLAVERY MOVEMENT WAS
25	28	EPIC ABOUT A GREEK WAR LEADER'S VOYAGE HOM
26	32	OLIVER, MICAWBER, PIP, AND GRADGRIND APEAR
27	11	LUCIFER IS ANOTHER NAME FOR SATAN
28	34	LEGENDARY KING ARTHUR PRESIDED OVER KNIGHT
29	38	IN BIBLE, SAMSON IS KNOWN FOR GREAT STRENG
30	51	JESUS WAS BETRAYED FOR 30 PIECES OF SILVER
31	54	E. DICKINSON, AMERICAN POET, LIVED IN SOLI
32	39	MORAL OF TORTOISE AND HARE: SLOW AND STEAD
33	46	IN CAIN AND ABEL STORY, JEALOUS YOUNG MAN
34	30	SHERLOCK HOLMES: MAIN CHARACTER IN HOUND O
35	48	EDGAR ALLAN POE WROTE POEMS "ANNABEL LEE"
36	35	IN ROMAN MYTHOLOGY, VENUS IS GODDESS OF LO
37	36	"WE HOLD THESE TRUTHS...,": DECLARATION OF
38	24	"WE, THE PEOPLE...": PREAMBLE TO THE CONST
39	44	THORNTON WILDER PLAY ABOUT TYPICAL NEW ENG
40	47	AESOP IS BEST KNOWN FOR WRITING FABLES
41	52	IN GREEK MYTHOLOGY, MULTITUDE OF EVILS ESC
42	80	HEMINGWAY WROTE "FOR WHOM THE BELL TOLLS",
43	59	JACK LONDON WROTE "CALL OF THE WILD" ABOUT
44	69	ROBERT FROST POEMS: "... PROMISES TO KEEP.
45	82	IN MELVILLE'S "MOBY DICK", CAPTAIN AHAB'S
46	61	LILLIPUT, COUNTRY OF LITTLE PEOPLE, PART O
47	42	"THE RED BADGE OF COURAGE": SOLDIER'S STRU
48	55	IN GREEK MYTHOLOGY, ATLAS HAS TO SUPPORT H
49	65	IN BIBLE, KING SOLOMON IS FAMOUS FOR HIS W

```
DENT SCHOOL DISTRICT                    8:01 FRIDAY, AUGUST 19, 1988    4
ANAGEMENT INFORMATION
ARCH AND EVALUATION

ERATURE TEST RANK ORDER RESULTS

LITERATURE ---------------------------------------------------------------

                              # STUDENTS            #          %
                            RECEIVING ITEM       CORRECT    CORRECT

 INTO ARK                         178              163        91.6
IS PEOPLE FROM EGYPT              165              149        90.3
IES                               159              145        91.2
                                  179              170        95.0
                                  165              143        86.7
A CHRISTMANS CAROL"               178              158        88.8
                                  165              149        90.3
CE IN WONDERLAND                  178              152        85.4
 TO GIVE TO POOR                  159              143        89.9
 MET PRINCE                       179              152        84.9
ND AND SURVIVED                   165              135        81.8
FROM PSALM 23                     178              151        84.8
ERLIN                             178              145        81.5
 AND RUNAWAY SLAVE                159              146        91.8
 OF CREATION                      165              133        80.6
OSOPHERS                          178              149        83.7
ORY OF DAVID & GOLIATH            179              142        79.3
R 20 YEARS AND AWOKE              165              130        78.8
F HOUSE OF USHER," ETC.           159              123        77.4
JULIUS CAESAR"                    179              157        87.7
H WE CALL A ROSE..."              178              140        78.7
ARS AGO OUR FATHERS..."           165              125        75.8
ATURE COULD NOT CONTROL           159              124        78.0
 "UNCLE TOM'S CABIN"              179              144        80.4
E IS THE "ODYSSEY"                159              124        78.0
 IN NOVELS OF DICKENS             159              121        76.1
                                  179              153        85.5
S OF ROUND TABLE                  165              125        75.8
TH                                159              117        73.6
 BY JUDAS                         178              113        63.5
TUDE, WROTE ABOUT DEATH           159              104        65.4
Y WINS THE RACE                   159              117        73.6
KILLS BROTHER                     165              114        69.1
F THE BASKERVILLES, ETC.          165              127        77.0
AND "RAVEN"                       178              122        68.5
VE                                178              134        75.3
 INDEPENDENCE                     159              118        74.2
ITUTION                           159              126        79.2
LAND VILLAGE: "OUR TOWN"          159              111        69.8
                                  179              123        68.7
APED FROM PANDORA'S BOX           165              109        66.1
 "THE SUN ALSO RISES"             159               86        54.1
 DOG IN YUKON                     178              112        62.9
...," "TWO ROADS..."              179              104        58.1
OBSESSION IS REVENGE              165               88        53.3
F "GULLIVER'S TRAVELS"            165              102        61.8
GGLE TO OVERCOME FEAR             165              119        72.1
EAVENS ON HIS SHOULDERS           179              116        64.8
ISDOM                             159               96        60.4
```

PROGRAM: SA$NAEP8

AUSTIN INDEPEN
DEPARTMENT OF M
OFFICE OF RESE

DISTRICTWIDE HISTORY & LIT

-- TEST-

National Rank	AISD Rank	ITEM DESCRIPTION
50	50	THE ILIAD IS AN EPIC POEM BY HOMER
51	64	IN BIBLICAL STORY, JONAH IS SWALLOWED BY B
52	40	IN MARK TWAIN NOVEL, TOM SAWYER IS KNOWN F
53	41	MIDAS WAS MYTHICAL KING WHOSE TOUCH TURNED
54	37	"THE SCARLET LETTER" IS NOVEL ABOUT WOMAN
55	19	"TO KILL A MOCKINGBIRD": 2 CHILDREN'S FATH
56	49	DICKENS' NOVEL TALE OF TWO CITIES OCCURS D
57	36	LONGFELLOW WROTE POEM ABOUT PAUL REVERE: "
58	51	FDR SAID: "THE ONLY THING WE HAVE TO FEAR.
59	31	IN ADDITION TO PLAYS, SHAKESPEARE ALSO WRO
60	68	YOUNG MAN WHO WASTES FORTUNE, COMES TO SEN
61	45	IN ROMAN MYTHOLOGY, MARS IS GOD OF WAR
62	76	W. CHURCHILL: "I HAVE NOTHING TO OFFER....
63	20	ARTHUR MILLER WROTE "THE CRUCIBLE" AND "DE
64	43	A "RAISIN IN THE SUN" IS ABOUT EXPERIENCES
65	63	JFK: "...ASK NOT WHAT YOUR COUNTRY CAN DO
66	62	IN GREEK MYTHOLOGY, JASON TRAVELED IN QUES
67	16	"THE GREAT GATSBY": NOVEL BY F. SCOTT FITZ
68	75	OEDIPUS: KILLED FATHER AND MARRIED MOTHER
69	88	"ACHILLES' HEEL" MEANS WEAK POINT IN STRON
70	60	IN GREEK MYTH, DAEDALUS & ICARUS MAKE WAX
71	73	ODYSSEUS IS MYTHICAL GREEK HERO WHO JOURNE
72	95	BYRON, KEATS, AND WORDSWORTH ARE CHIEFLY K
73	53	SHAKESPEARE'S "JULIUS CAESAR" IS ABOUT THE
74	56	DON QUIXOTE WAS SPANISH KNIGHT WHO ATTACKE
75	103	MACBETH: PLAY ABOUT MAN LED TO MURDER BY A
76	89	"TO EVERY THING THERE IS A SEASON,...": TH
77	99	IN BEOWULF, THE HERO BATTLES WITH THE MONS
78	78	ABRAHAM LINCOLN: "WITH MALICE TOWARD NONE,
79	92	T.S. ELIOT WROTE "THE WASTE LAND", "PRUFRO
80	79	IN GREEK MYTH, TROJAN WAR STARTS BECAUSE P
81	87	MAXIMS FROM FRANKLIN'S "POOR RICHARD'S ALM
82	58	POEMS WRITTEN BY NATHANIEL HAWTHORNE
83	67	"THE OLD MAN AND THE SEA": NOVEL ABOUT MAN
84	74	"WALDEN" BY HENRY DAVID THOREAU ABOUT SIMP
85	108	MILTON'S "PARADISE LOST": REBELLION OF SAT
86	81	LEAVES OF GRASS BY WALT WHITMAN INCLUDES "
87	71	STEINBECK'S THE GRAPES OF WRATH: FAMILY MI
88	77	GREEK PLAY ANTIGONE: WOMAN DEFIES KING TO
89	98	"THINGS FALL APART;..." FROM YEATS' "THE S
90	85	IN GREEK MYTH, PROMETHEUS IS CHAINED TO A
91	101	AUSTEN'S PRIDE AND PREJUDICE: THE BENNET D
92	100	"WUTHERING HEIGHTS": HEATHCLIFF'S OBSESSIV
93	105	G.B. SHAW WROTE "PYGMALION", "ARMS AND THE
94	94	IN THE BIBLE, JOB IS KNOWN FOR HIS PATIENC
95	91	WALTER MITTY HAD DULL EXISTENCE BUT EXCITI
96	66	LANGSTON HUGHES, POET OF HARLEM RENAISSANC
97	93	GEOFFREY CHAUCER WROTE "THE CANTERBURY TAL
98	84	HERMAN MELVILLE WROTE "BILLY BUDD", "BENIT

```
DENT SCHOOL DISTRICT                          8:01 FRIDAY, AUGUST 19, 1988     5
ANAGEMENT INFORMATION
ARCH AND EVALUATION

ERATURE TEST RANK ORDER RESULTS

LITERATURE -----------------------------------------------------------------

                                    # STUDENTS           #            %
                                 RECEIVING ITEM      CORRECT      CORRECT

                                       179              121          67.6
IG FISH                                179              109          60.9
OR CLEVER WAYS OF WORK                 179              131          73.2
 OBJECTS TO GOLD                       159              115          72.3
WHO WAS UNFAITHFUL                     178              131          73.6
ER DEFENDS BLACK MAN                   178              145          81.5
URING FRENCH REVOLUTION                165              112          67.9
ONE IF BY LAND,...."                   165               81          49.1
...," "12-7-41...INFAMY"               179              120          67.0
TE SONNETS                             179              137          76.5
SES LIKE PRODIGAL SON                  179              105          58.7
                                       179              124          69.3
" "...IRON CURTAIN..."                 178               99          55.6
ATH OF A SALESMAN"                     165              134          81.2
 OF BLACK FAMILY                       165              116          70.3
FOR YOU,...."                          165              101          61.2
T OF THE GOLDEN FLEECE                 159               98          61.6
GERALD                                 179              149          83.2
IN ANCIENT GREEK PLAY                  179              100          55.9
G PERSON                               178               86          48.3
WINGS, FLY CLOSE TO SUN                159              100          62.9
YS HOME AFTER TROJAN WAR               165               94          57.0
NOWN AS POETS                          165               75          45.5
 FATE OF HIS ASSASSINS                 178              117          65.7
D WINDMILLS                            159              101          63.5
MBITION TO BE KING                     165               64          38.8
E BIBLE                                159               76          47.8
TER GRENDEL                            159               64          40.3
 W/ CHARITY FOR ALL..."                178               97          54.5
CK", "THE HOLLOW MEN"                  165               76          46.1
ARIS KIDNAPPED HELEN                   178               97          54.5
ANACK",                                179               87          48.6
                                       178              113          63.5
 WHO BATTLES GREAT FISH                179              108          60.3
LIFYING ONE'S LIFE                     165               94          57.0
AN, FALL OF ADAM & EVE                 178               60          33.7
I CELEBRATE MYSELF,...."               178               95          53.4
GRATES TO CALIFORNIA                   179              103          57.5
HONOR HER DEAD BROTHER                 178               99          55.6
ECOND COMING"                          178               76          42.7
ROCK FOR STEALING FIRE                 178               88          49.4
AUGHTERS FIND HUSBANDS                 179               71          39.7
E LOVE OF CATHERINE                    179               72          40.2
 MAN", AND "SAINT JOAN"                179               62          34.6
E DURING SUFFERING                     178               81          45.5
NG FANTASY LIFE                        179               83          46.4
E: "HOLD FAST TO DREAMS"               159               96          60.4
ES"                                    179               82          45.8
D CERENO", ETC.                        179               89          49.7
```

PROGRAM: SA$NAEP8

AUSTIN INDEPEND
DEPARTMENT OF MAN
OFFICE OF RESEAR

DISTRICTWIDE HISTORY & LITER

-- TEST=L

National Rank	AISD Rank	ITEM DESCRIPTION
99	90	1984: DICTATORSHIP WATCHING EVERYONE TO STA
100	70	LORD OF THE FLIES: STRANDED CHILDREN TRY &
101	83	HERMAN MELVILLE AND JOSEPH CONRAD WROTE NOV
102	96	BIBLE: SODOM & GOMORRAH ARE DESTROYED DUE T
103	106	THE SOUTH WAS SETTING FOR MANY OF WILLIAM F
104	97	DANTE'S DIVINE COMEDY: JOURNEY THROUGH HELL,
105	107	RICHARD WRIGHT WROTE "NATIVE SON" AND "BLAC
106	109	D.H. LAWRENCE WROTE "ROCKING HORSE WINNER"
107	115	WILLA CATHER WROTE "MY ANTONIA", "O PIONEER
108	72	TENNESSEE WILLIAMS WROTE STREETCAR NAMED DE
109	104	HEMINGWAY WROTE "IN ANOTHER COUNTRY", "SHOR
110	116	HARDY WROTE RETURN OF THE NATIVE, TESS OF T
111	114	CATCHER IN THE RYE: BOY OF 16 EXPELLED FROM
112	117	HENRY JAMES WROTE "DAISY MILLER" AND "PORTR
113	102	IBSEN WROTE HEDDA GABLER, DOLL'S HOUSE, AN
114	113	JOSEPH CONRAD WROTE HEART OF DARKNESS, LORD
115	111	ELLISON'S INVISIBLE MAN: YOUNG MAN FROM SOU
116	110	DOSTOEVSKY WROTE CRIME AND PUNISHMENT AND T
117	119	JAMES JOYCE WROTE ULYSSES, PORTRAIT OF THE
118	112	DE TOCQUEVILLE: EUROPEAN WHO WROTE "DEMOCRA
119	118	EUDORA WELTY & FLANNERY O'CONNOR KNOWN FOR
120	121	WILLIAM BLAKE REFERS TO A TIGER IN POEM "...
121	120	BUNYAN'S PILGRIM'S PROGRESS: TEMPTATIONS TH

```
NT SCHOOL DISTRICT                         8:01 FRIDAY, AUGUST 19, 1988    6
NAGEMENT INFORMATION
RCH AND EVALUATION

RATURE TEST RANK ORDER RESULTS

TERATURE ----------------------------------------------------------------------

                                # STUDENTS            #            %
                             RECEIVING ITEM        CORRECT      CORRECT

IP OUT INDIVIDUALITY               159               76           47.8
AIL TO BE CIVILIZED                179              104           58.1
LS ABOUT THE SEA                   178               91           51.1
) WICKED INHABITANTS               165               74           44.8
ULKNER'S NOVELS                    159               54           34.0
 PURGATORY, HEAVEN                 165               73           44.2
: BOY"                             159               54           34.0
ND "SONS AND LOVERS"               159               52           32.7
SI" ETC.                           165               42           25.5
SIRE, GLASS MENAGERIE              178              102           57.3
 HAPPY LIFE...", ETC.              165               63           38.2
IE D'URBERVILLES, ETC.             159               37           23.3
SCHOOL, GOES TO NYC                178               46           25.8
IT OF A LADY"                      179               37           20.7
NEMY OF THE PEOPLE                 159               63           39.6
 JIM, SECRET SHARER                165               43           26.1
H MOVES TO HARLEM                  178               55           30.9
IE BROTHERS KARAMAZOV              165               51           30.9
RTIST..., ARABY, ETC.              179               32           17.9
Y IN AMERICA"                      159               42           26.4
TORIES SET IN SOUTH                178               33           18.5
BURNING BRIGHT..."                 159               21           13.2
T CHRISTIANS FACE                  178               24           13.5
```

ATTACHMENT 9

ITEMS TESTED ON THE U.S. HISTORY ASSESSMENT

Weighted Percent Correct in Descending Order by National Performance

Description	*Total*
Thomas Edison invented the light bulb	95.2
Find Soviet Union on map of Europe	92.1
Alexander Graham Bell invented the telephone	91.1
George Washington was president between 1780–1800	87.9
Find Italy on map of Europe	87.7
The Underground Railroad was secret network to help slaves escape	87.5
Adolf Hitler was leader of Germany during World War II	87.4
Thomas Jefferson was main author of Declaration of Independence	87.4
The assembly line was introduced by US auto industry	87.2
Find area of 13 original states on map of US	84.8
Ku Klux Klan used violence to oppose minorities	83.9
Harriet Tubman was leader in helping slaves escape	83.8
Find Rocky Mountains on map of US	81.3
Guarantee of free speech and religion found in Bill of Rights	81.3
Japanese attack on Pearl Harbor led US into World War II	80.0
US dropped first atomic bomb on Japan in World War II	79.9
Washington was commander of American army in Revolution	79.2
Churchill was prime minister of Great Britain in World War II	78.1
TV became new feature in American homes after 1950	78.1
Watergate led to resignation of Richard Nixon	77.4
Women worked in factories during World War II	77.3
Jamestown was first permanent English colony	76.1
Find West Germany on map of Europe	76.1
Lindbergh made first solo flight across Atlantic	76.1
Nazi decimation of Jewish people in Europe called the Holocaust	75.8
The Great Depression was a period of mass unemployment	75.1
"Prohibition" refers to ban on sale and consumption of liquor	74.6
The Great Depression occurred between 1900–1950	72.2
Vietnam War had strongest citizen protest movements	72.1
Civil rights movement of 1960s focused on equality for minorities	71.7
Discovery of gold in California was a cause of westward movement	71.3
Find area on US map that fought for independence from Mexico	71.0
Invention of cotton gin stimulated plantation economy in South	70.7
Germany and Japan were main US enemies in World War II	70.7
World War II ended between 1943–1947	70.7
Indians were put on reservations after Civil War	70.5
Find Mississippi River on map of US	70.3
Find Great Britain on map of Europe	70.2
Most people in English colonies of North America worked as farmers	69.7
"Secession" refers to withdrawal of Southern states from Union	69.7
Susan B. Anthony was a leader of women's suffrage movement	68.9
Columbus discovered the New World before 1750	68.1
Abraham Lincoln wrote the Emancipation Proclamation	68.0
The Declaration of Independence was signed between 1750–1800	67.8
The Declaration of Independence marks colonies' break with England	67.6
Stamp Act of 1765 was attempt by English to tax colonies	67.3
US policy after World War II was to provide aid to Europe	67.2
Find France on map of Europe	65.8
US provoked to enter World War I by German submarine attacks	64.6
Watergate occurred after 1950	64.5
Brown decision of 1954 ruled school segregation unconstitutional	63.7
Alexander Hamilton known for urging strong national government	63.5
Find area of the Confederacy on map of US	63.4
Sputnik was the first man-made satellite	62.7
Plessy vs. *Ferguson* decision approved racial segregation	61.4
The southwestern US was explored and settled by Spain	61.0
United States Constitution was written between 1750–1800	60.9
East Coast of US explored and settled mainly by England	60.6
Woodrow Wilson appealed for American entry into League of Nations	60.2
Japan bombed Pearl Harbor between 1939–1943	60.0
"Checks and balances" divides power among branches of federal government	59.9
Heavy immigration in 1970s–80s came from SE Asia and Latin America	59.5
Articles of Confederation were first Constitution during Revolution	59.4
Representative government existed in Colonies	58.0
John D. Rockefeller formed Standard Oil	57.8
US foreign policy after World War II was containment of Communism	57.7
World War II was between 1900–1950	57.3
Booker T. Washington was major black leader before World War I	57.1
Find area of map of US purchased from France	57.1
"Prohibition" occurred between 1900–1950	56.4
Before 1800 most immigrants came from northern and western Europe	56.3
Eisenhower was president between 1946–1963	55.6
Israel has never been invaded by Soviet Union	55.4
Japanese-Americans were forced into camps during World War II	55.2
Nixon established relations with the People's Republic of China	55.1
A cause of population movement in 1800s was reasonable land prices	54.7
Social Security system was introduced during the New Deal	54.7
US gave military support to South Korea with United Nations	53.9
Dust Bowl of 1930s fostered movement to California	53.8
Amendment of 1920 granted women right to vote	53.7
Stalin was leader of Soviet Union during World War II	53.6
Populist Party advocated government support for farmers	52.8
New Deal identified with changes in social, economic policies	52.3
Franklin Roosevelt was president between 1929–1946	52.0
Before Civil War, nation debated spread of slavery	51.6
"Give me liberty or give me death"—Patrick Henry	51.1
"Laissez-faire" means minimal government regulation of economy	51.0
Monroe Doctrine said Europe should not gain land in western hemisphere	50.9
Washington's Farewell Address warned against foreign alliances	50.3
Canada and Mississippi valley first explored by France	50.3
Samuel Gompers was first president of AFL	49.8

ITEMS TESTED ON THE U.S. HISTORY ASSESSMENT *(Continued)*

Description	*Total*
Martin Luther King, Jr., rose to prominence in Montgomery bus boycott	48.9
US dropped first atomic bomb between 1943–1947	48.6
Valley Forge was lowest point in Revolutionary War	47.9
Women's right to vote was not guaranteed after Civil War	47.3
An issue in War of 1812 was Great Britain's interference with shipping	47.0
Andrew Carnegie was associated with development of steel industry	46.9
Thomas Jefferson was president between 1800–1820	45.6
The Constitution divides powers between states and federal government	43.8
Missouri Compromise admitted Maine as free, Missouri as slave state	43.0
Woodrow Wilson was president between 1912–1929	42.9
Senator Joseph McCarthy involved in controversy about Communism	42.6
"Nullification" related to states' rights	42.4
Hoover, Franklin Roosevelt were presidents during Depression	41.1
Jane Addams founded settlement houses to help the poor	41.0
"Reconstruction" occurred between 1850–1900	40.2
The Federalist advocated adoption of Constitution	40.1
Dred Scott decision: slave who moved to free state was not free	39.5
D-Day occurred between 1943–1947	39.5
Renaissance was characterized by cultural, technological advances	39.3
Paine's *Common Sense* argues for colonial independence	38.3
Emancipation Proclamation freed slaves in Confederacy	38.2
Union membership grew in the 1930s because of new laws	38.2
Jamestown founded before 1750	38.0
Restrictions on immigration were not part of New Deal	37.8
"Three-fifths compromise" in Constitution defined status of slaves	37.7
Immigration from southern and eastern Europe grew, 1890–1910	37.6
Immigration restriction in 1921, 1924 aimed at SE Europeans	37.3
Scopes trial was about teaching evolution	37.2
Upton Sinclair, Lincoln Steffens, Ida Tarbell known as muckrakers	37.1
Theodore Roosevelt was president between 1895–1912	36.9
Articles of Confederation failed to provide adequate taxing power	36.8
Find region on US map acquired from Mexico in war	36.2
Religious toleration in colonies due to common interest of many groups	36.0
Jonas Salk invented polio vaccine	34.3
Spanish-American War made US an international power	33.0
American foreign policy after World War I known as isolationist	32.3
The Civil War occurred between 1850–1900	32.2
US foreign policy in early 1900s: "Speak softly, carry a big stick"	31.6
Purpose of Jim Crow laws was to enforce racial segregation	30.7
Magna Carta is foundation of British parliamentary system	30.6
Andrew Jackson was president between 1820–1840	29.9
Reformation led to establishment of Protestant sects	29.8
The United Nations was founded between 1943–1947	25.9
The Seneca Falls Declaration concerned women's rights	25.8
Abraham Lincoln was president between 1860–1880	24.7
Lyndon Johnson's term included Medicare and Voting Rights Act	23.9
Betty Friedan and Gloria Steinem led women's movement in 1970s	22.8
"Progressive movement" refers to reforms before World War I	22.6
"Reconstruction" refers to readmission of Confederate states	21.4
John Winthrop and the Puritans founded a colony in Boston	19.5

Note: The descriptions above are paraphrases of the information tested. The answers are not identical to the material on the test. In some instances, for reasons of space, the information is incomplete as presented here. Wherever a bloc of time is in the answer, students were presented with several choices on a timeline; some timelines were divided into half-centuries, some into twenty-year periods, and some into even shorter spans.

ITEMS TESTED ON THE LITERATURE ASSESSMENT

Weighted Percent Correct in Descending Order by National Performance

Description	*Total*
In Bible, Noah gathered pairs of creatures into ark	94.0
In Bible, Moses got 10 commandments, led his people from Egypt	92.3
Romeo and Juliet hindered by feuding families	89.7
Martin Luther King, Jr.: "I have a dream . . ."	88.1
Hamlet: "To be or not to be . . ." (5 lines)	87.8
Ebenezer Scrooge was stingy character in "A Christmas Carol"	87.2
In Greek mythology, ruler of gods is Zeus	86.7
White Rabbit, March Hare, etc., are in *Alice's Adventures in Wonderland*	86.1
Robin Hood is known for stealing from rich to give to poor	85.7
Cinderella's rags turned into gown and she met prince	85.1
Robinson Crusoe was shipwrecked on an island and survived	83.6
"The Lord is my shepherd . . ." (3 lines) is from Psalm 23	82.4
The magician who advised King Arthur was Merlin	80.5
Adventures of Huckleberry Finn is about orphan boy and runaway slave	80.5
In Bible, Book of Genesis includes account of creation	79.5
Plato and Aristotle are best known as philosophers	79.0
Victory of unknown team parallels Bible story of David and Goliath	78.0
"Rip van Winkle" is about man who slept for 20 years and awoke	76.3
E. A. Poe wrote "Pit and Pendulum," "Fall of House of Usher," etc.	75.2
"Friends, Romans, countrymen . . ." is from *Julius Caesar*	74.9
Juliet said, "What's in a name? That which we call a rose . . ."	74.2
Gettysburg Address: "Four score and seven years ago our fathers . . . ," etc.	73.9
Mary Shelley's novel *Frankenstein* is about scientist who created creature he could not control	73.8
Novel that helped the antislavery movement was *Uncle Tom's Cabin*	73.4
Epic about a Greek war leader's voyage home is the *Odyssey*	73.0
Oliver, Micawber, Pip, and Gradgrind appear in novels of Dickens	72.7
Lucifer is another name for Satan	72.3
Legendary King Arthur presided over knights of Round Table	72.0
In Bible, Samson is known for great strength	71.8
Jesus was betrayed for 30 pieces of silver by Judas	69.5
Emily Dickinson was American poet who lived in solitude, wrote often about death (lines from 2 poems)	69.3
Moral of tortoise and hare: slow and steady wins the race	67.7
In Cain and Abel story, jealous young man kills brother	67.3
Sherlock Holmes is main character in *Hound of the Baskervilles*, etc.	67.2
Edgar Allan Poe wrote poems "Annabel Lee" and "Raven"	67.0
In Roman mythology, Venus is goddess of love	66.5
"We hold these truths to be self-evident, that all men are created equal . . . ," etc., is in the Declaration of Independence	65.7
"We, the people of the United States, in order to form a more perfect union," is from the Preamble to the Constitution	65.7
Play by Thornton Wilder about typical New England village is *Our Town*	65.6
Aesop is best known for writing fables	65.3
In Greek mythology, multitude of evils escaped from Pandora's box	64.0
Hemingway wrote *For Whom the Bell Tolls, The Sun Also Rises*	63.2
Jack London wrote *Call of the Wild* about dog in Yukon	62.5
Robert Frost wrote poems ("But I have promises to keep . . . ," "Two roads diverged in a wood . . .")	62.5
In Melville's *Moby-Dick*, Captain Ahab's obsession is revenge	61.8
Lilliput, country of little people, is part of *Gulliver's Travels*	61.7
The Red Badge of Courage is about young soldier's struggle to overcome his fear in Civil War	61.6
In Greek mythology, Atlas has to support heavens on his shoulders	61.1
In Bible, King Solomon is famous for his wisdom	61.0
The *Iliad* is an epic poem by Homer	60.6
In biblical story, Jonah is swallowed by big fish	60.2
In Mark Twain novel, Tom Sawyer is known for clever ways of avoiding work, trouble	59.8
Midas was mythical king whose touch turned objects to gold	59.5
The Scarlet Letter is novel about woman who was unfaithful	59.4
To Kill a Mockingbird is novel about 2 children affected by community conflict when father defends black man	59.1
Dickens's novel *Tale of Two Cities* occurs during French Revolution	59.0
Longfellow wrote poem about Paul Revere that includes "One if by land, / Two if by sea"	58.8
Franklin Roosevelt said, "The only thing we have to fear is fear itself" and "Yesterday, December 7, 1941,—a date which will live in infamy . . ."	57.2
In addition to plays, Shakespeare also wrote sonnets	57.0
Young man who wastes fortune, comes to his senses is like prodigal son	56.5
In Roman mythology, Mars is god of war	56.2
Winston Churchill said, "I have nothing to offer, but blood, toil, tears, and sweat" and "From Stettin in the Baltic to Trieste in the Adriatic, an Iron Curtain has descended across the continent"	55.7
Arthur Miller wrote *The Crucible* and *Death of a Salesman*	53.7
A Raisin in the Sun is about experiences of black family	53.2
John F. Kennedy: "And so, my fellow Americans, ask not what your country can do for you; ask what you can do for your country"	52.7
In Greek mythology, Jason traveled in quest of the Golden Fleece	52.5
The Great Gatsby is novel by F. Scott Fitzgerald about pursuit of wealth and status in 1920s	51.7
Character in ancient Greek play who unknowingly killed father and married his mother is Oedipus	51.7
"Achilles' heel" means weak point in strong person	51.5
In Greek mythology, Daedalus and Icarus make wax wings, fly too close to sun, fall into sea	50.2

BEST COPY AVAILABLE

ITEMS TESTED ON THE LITERATURE ASSESSMENT *(Continued)*

Description	*Total*
Odysseus is mythical Greek hero who journeys home after Trojan War	49.8
Byron, Keats, and Wordsworth are chiefly known as poets	48.1
Shakespeare's play *Julius Caesar* is about the fate of his assassins	48.0
Don Quixote was Spanish knight who attacked windmills, thinking they were giants	47.9
Macbeth is play about man whose ambition to be king led him to murder	47.3
"To every thing there is a season, and a time to every purpose under the heaven: a time to be born, and a time to die": the Bible	46.7
In *Beowulf*, the hero battles with the monster Grendel	45.8
Abraham Lincoln: "With malice toward none, with charity for all . . ."	45.7
T. S. Eliot wrote *The Waste Land*, "The Love Song of J. Alfred Prufrock," and "The Hollow Men"	45.4
In Greek mythology, Trojan War starts because Paris kidnapped Helen	45.4
"A penny saved is a penny earned" and "A small leak will sink a great ship" are maxims from Franklin's *Poor Richard's Almanack*	43.6
"The Minister's Black Veil," "Young Goodman Brown," "Rappaccini's Daughter" were written by Nathaniel Hawthorne	43.3
The Old Man and the Sea is novel about man who battles great fish	43.0
Walden by Henry David Thoreau is about simplifying one's life	42.9
John Milton's *Paradise Lost* is about rebellion of Satan, fall of Adam and Eve	41.2
Walt Whitman wrote *Leaves of Grass*, which includes line, "I celebrate myself, and sing myself"	40.2
Steinbeck's *The Grapes of Wrath* is about family that migrates from Dust Bowl to California	39.7
Antigone is Greek play about woman who defies king to honor her dead brother	39.1
"Things fall apart; / The center cannot hold . . . ," etc., is from "The Second Coming" by William Butler Yeats	38.9
In Greek mythology, Prometheus is chained to a rock for stealing fire	38.5
In Jane Austen's *Pride and Prejudice*, the Bennet daughters find husbands	38.0
Wuthering Heights is about Heathcliff's obsessive love of Catherine	37.9
George Bernard Shaw wrote *Pygmalion*, *Arms and the Man*, and *Saint Joan*	37.5
In the Bible, Job is known for his patience during suffering	37.2
Walter Mitty had dull existence but exciting fantasy life	36.6
Langston Hughes was poet of Harlem Renaissance who wrote "Hold fast to dreams," etc.	36.2
Geoffrey Chaucer wrote *The Canterbury Tales*	36.1
Herman Melville wrote *Billy Budd*, "Benito Cereno," and "Bartleby the Scrivener"	35.9
1984 is novel about dictatorship that watches everyone to stamp out individuality	35.5
Lord of the Flies is novel about children stranded on island who try and fail to lead civilized life	35.3
Herman Melville and Joseph Conrad are authors of novels about the sea	34.7
In the Bible, the cities of Sodom and Gomorrah are destroyed because of wicked inhabitants	33.4
The South was setting for many of William Faulkner's novels	33.3
Dante's *Divine Comedy* is about journey through Hell, Purgatory, Heaven	32.8
Richard Wright wrote *Native Son* (about black life in Chicago) and *Black Boy*	32.3
D. H. Lawrence is English author who wrote "The Rocking Horse Winner" and *Sons and Lovers*	28.7
Willa Cather wrote about settling of West in *My Ántonia*, *O Pioneers!*, and *Death Comes for the Archbishop*	28.2
Tennessee Williams wrote *A Streetcar Named Desire* and *The Glass Menagerie*	27.6
Ernest Hemingway wrote "In Another Country," "The Short Happy Life of Francis Macomber," and "The Killers"	27.3
Thomas Hardy wrote *The Return of the Native*, *Tess of the D'Urbervilles*, *The Mayor of Casterbridge*	24.4
Catcher in the Rye is novel about 16-year-old boy who is expelled from school, goes to NYC for weekend	22.5
Henry James wrote about American/European life in *Daisy Miller* and *Portrait of a Lady*	21.9
Henrik Ibsen wrote *Hedda Gabler*, *A Doll's House*, *An Enemy of the People*	20.3
Joseph Conrad wrote *The Heart of Darkness*, *Lord Jim*, and *The Secret Sharer*	19.3
Ralph Ellison's *Invisible Man* is about young man who grows up in South, moves to Harlem	18.3
Dostoevsky wrote *Crime and Punishment* and *The Brothers Karamazov*	17.1
James Joyce is Irish author who wrote *Ulysses*, *A Portrait of the Artist As a Young Man*, "Araby" and "Eveline"	15.6
De Tocqueville was European who traveled in US and wrote *Democracy in America*	15.5
Eudora Welty and Flannery O'Connor are known for stories set in South	14.4
William Blake refers to a tiger in poem, "——! ——! burning bright / In the forests of the night, / What immortal hand or eye / Could frame thy fearful symmetry?"	13.6
John Bunyan's novel *Pilgrim's Progress* is about the temptations that Christians face in life	13.4

Note: These descriptions are paraphrases of the information tested. In most instances, for reasons of space, the information is an abbreviated and sometimes incomplete version of the item on the assessment.

ATTACHMENT 10

June 1, 1988

Name/Title
Position
Affiliation

Dear

I would like to request that you serve in the capacity of reviewer for the Austin Independent School District's Content in the Curriculum project.

The Content in the Curriculum project evolved in response to the growing national concern that students may lack basic knowledge in U.S. history and literature. In the spring of 1986, the National Assessment of Educational Progress (NAEP) tested a nationally representative sample of 17-year-olds to assess their basic knowledge of U.S. history and literature. The results of this assessment were published in a best-selling book by Diane Ravitch and Chester Finn entitled What Do Our 17-Year-Olds Know?

In AISD, we are likewise interested in what our students know, but we are also concerned with several other fundamental questions: (1) What is important for us to teach our children? (2) Are these important things in our curriculum? and (3) Are we teaching these important things to our children?

In order to address the second and third questions, it is necessary to obtain some guidance and direction in regard to the first question. To do this, we are asking persons with a stake in humanities instruction to review the NAEP item specifications and to rate them as to their importance for AISD students. The item specifications for either U.S. history or literature, along with additional information and instructions for rating, are enclosed.

Your assistance as a reviewer will help ensure that AISD's U.S. history and literature curricula contain the content our young people need to know to participate fully in our complex modern society. If you are unable to take on this task, please get in touch with David Wilkinson at 458-1227. I hope, however, that you will be able to assist the Austin schools by serving as a reviewer.

Sincerely,

John Ellis
Superintendent of Schools

ATTACHMENT 11

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

CONTENT IN THE CURRICULUM

Review of NAEP Item Specifications

BACKGROUND INFORMATION

In 1986, 7,812 17-year-old students across the nation were tested on their knowledge of U.S. history and literature. Conducted by the National Assessment of Educational Progress (NAEP), one of the nation's leading testing agencies, the test was the first of its kind. The students were drawn from public and private schools, and the sample was representative of different ethnicities, both sexes, and all regions of the United States.

Although the test was intended to be a measure of basic knowledge in the two areas, students could correctly answer only 55% of the history questions and 52% of the literature questions.

Diane Ravitch and Chester E. Finn, Jr., authors of the best-selling book *What Do Our 17-Year-Olds Know?*, which presents the results of the assessment, have characterized these results as "a shameful level of performance."

In the Austin Independent School District, the implications of the national results dovetail with a local concern with the amount of content being taught and the balance between skills and knowledge in the curriculum. The Superintendent and his cabinet have raised the question, "How does AISD determine what the content of instruction ought to be?" To begin to answer this question, the Office of Research and Evaluation has undertaken an evaluation of the content in American history and literature.

One key feature of this evaluation centers around the question, "What are the important things that we should teach our children?" Taking as a starting point the domain of content defined by the National Assessment, a group of persons with a stake in instruction in AISD (teachers, instructional coordinators, subject area professors from Austin universities, etc.) is being asked to review the NAEP item specifications and rate them as to their importance for AISD students.

These ratings will serve to inform the remainder of the evaluation. Phase two of the evaluation centers around the question, "Are the important things in our curriculum?" A review will be conducted of the Texas essential elements, the scope and sequence of textbooks, and AISD course outlines for evidence that the important content is included in the curriculum. Phase three involves an analysis of teachers' tests under the assumption that teachers test what is taught. The final phase of the evaluation would be testing AISD eleventh graders so that AISD's performance could be compared with the performance of students nationwide.

INSTRUCTIONS FOR RATING

You will be rating the objectives and subobjectives of one of the areas of the NAEP history and literature assessment, either history or literature. The NAEP learning objectives were created by two national committees, one for history and one for literature. The objectives were the basis for the 121 literature and 141 history items administered in the assessment. The objectives represent the consensus each committee arrived at about what should be tested. However, we are interested in what you think is important that AISD students be taught.

In the blank to the left of each objective or subobjective, write the number corresponding to the rating you select, according to the following scale.

Essential	Important	Somewhat Important	Unimportant
1	2	3	4

Please feel free to write comments about the basis for your ratings.

If there are other areas not addressed by the objectives which you think are important for students to know, please list them at the end. Write on the backs of the sheets with the objectives or attach additional pages.

When you have completed your rating, please use the enclosed stamped, self-addressed envelope to return the sheets to:

David Wilkinson/ORE
6100 Guadalupe
Austin, TX 78752

Please return your ratings by June 10, 1988.

If you have any questions about the rating task or about the evaluation, please call David Wilkinson at 458-1227.

THANK YOU VERY MUCH FOR YOUR TIME AND YOUR HELP.

ATTACHMENT 12

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

THE NAEP LEARNING OBJECTIVES

HISTORY

______ I. Exploration and Colonization: up to 1763

______ A. Exploration
______ 1. Factors contributing to exploration (e.g., desire for wealth, technological advances)
______ 2. Major countries and explorers involved (e.g., Spain, France, England)

B. Colonization
______ 1. Factors contributing to colonization (e.g., religious, economic, and social issues)
______ 2. The first colonies (e.g., governments based on English models, leaders of original colonies, and relations with Native American cultures)
______ 3. Aspects of colonial life (e.g., diversity of religious and ethnic groups, origins of slavery, types of economy. and immigration)

______ II. The Revolutionary War and the New Republic: 1763-1815

______ A. The Revolutionary War
______ 1. Factors contributing to the Revolutionary War (e.g., English exploitation of the colonies and emergence of an American Society)
______ 2. Documents (e.g., Declaration of Independence, the Articles of Confederation, and Common Sense)
______ 3. Importance of American leadership (e.g., George Washington, Thomas Paine, John Adams, Thomas Jefferson, Benjamin Franklin)
______ 4. Events and aspects of the war (e.g., political parties, leaders)

______ B. Establishing the new nation
______ 1. The Constitution (e.g., major compromises, structure, and steps leading to ratification)
______ 2. Forming the new government (e.g., political parties, leaders)
______ 3. Expansion (e.g., migration, the Louisiana Purchase, trade with Europe, difficulties maintaining neutrality, and the War of 1812)

______ III. Nationhood, Sectionalism, and the Civil War: 1815-1877

______ A. Economic and social change (e.g., growth of cities, industrialization, transportation)

______ B. Jacksonian democracy (e.g., political parties, expanding the franchise, treatment of Native Americans)

______ C. Expansion of slavery (e.g., Missouri Compromise, plantation economy, and abolitionists)

______ D. The Civil War
______ 1. Federal powers versus states' rights (e.g., nullification)
______ 2. Factors leading up the the Civil War (e.g., slavery, economic differences between North and South, and secession versus preservation of the Union)
______ 3. Abraham Lincoln (e.g., Emancipation Proclamation, Gettysberg Address)
______ 4. Effects of the Civil War (e.g., growth of North; destruction of South; Reconstruction; passage of Thirteenth Fourteenth, and Fifteenth Amendments--abolition of slavery, due process and equal protection, and the right to vote)

______ IV. Territorial Expansion, the Rise of Modern America, and World War I: 1877-1920

______ A. Territorial expansion
______ 1. Western expansion (e.g., territories involved, improved transportation, farm protests, Indian Wars, and the reservation system)
______ 2. The Spanish-American War (e.g., territorial acquisition, United States becomes world power)

______ B. The rise of modern America
______ 1. Big business (e.g., leaders, new production techniques, and monopolies)
______ 2. Labor unions (e.g., working conditions, American Federation of Labor, problems in organizing unions)
______ 3. Progressive Era and reform legislation (e.g., populism, Theodore Roosevelt, muckraking, "trust busting," and conservation)
______ 4. Immigration (e.g., shift in patterns, efforts to restrict--National Origins Act)
______ 5. Segregation and failure to achieve equality for blacks (e.g., Plessy v. Ferguson--separate but equal, Jim Crow laws, NAACP)

C. The First World War
   1. Reasons for United States entry (e.g., rights of neutrality and submarine warfare)
   2. Characteristics of the war (e.g., countries involved, leaders--Woodrow Wilson)
   3. Events and effects (Treaty of Versailles, League of Nations, substantial American contribution, isolations mood after war)

D. Women's vote--Nineteenth Amendment (e.g., early advocates--Susan B. Anthony, Elizabeth Cady Stanton, Seneca Falls Convention)

V. The Great Depression, the New Deal, and World War II: 1920-1945

A. The 1920's (e.g., temperance movement and prohibition, inventions, Scopes trial)

B. Causes and characteristics of the Great Depression (e.g., stock market crash, collapse of economy, Dust Bowl)

C. Franklin D. Roosevelt and the New Deal (e.g., changes in role of government, gains for labor, agricultural price supports, Social Security)

D. World War II
   1. Factors leading up to United States involvement in the war (e.g., rise of totalitarianism, United States initially neutral, entry after Japanese attack on Pearl Harbor)
   2. Characteristics of the war (e.g., global nature with European and Pacific theaters; the Holocaust; leaders--Churchill, Stalin, Hitler, and Roosevelt; effect of war on roles of women and minorities; Japanese relocation camps; relationships with Allies--Yalta)
   3. End of the Second World Was (e.g., Roosevelt dies and is succeeded by Truman; United States uses atomic bomb to end war with Japan)
   4. The United Nations (e.g., purpose)

VI. Post-World War II: 1945 to Present

A. The Cold War (e.g., containment of communism, beginnings of arms race, Truman Doctrine, Marshall Plan, NATO, McCarthyism, communist expansion in Europe)

B. Korean conflict (e.g., U.N. forces, MacArthur versus Truman)

______ C. Postwar prosperity (e.g., demand for consumer goods; the baby boom; the growth of suburbs; inventions and discoveries; Sputnik begins space race)

______ D. The 1960's
- ______ 1. President Kennedy and the New Frontier (e.g., space program, Cuban missle crisis, Peace Corps, assassination)
- ______ 2. President Lyndon Johnson and the Great Society (e.g., increased social legislation and government spending)
- ______ 3. Civil rights movement (e.g., Brown v. Board of Education and beyond, Martin Luther King, Jr., civil rights legislation)
- ______ 4. Vietnam (e.g., growing unpopularity of war, protest movement)

______ E. The 1970's
- ______ 1. Opening to China
- ______ 2. United States withdrawal from Vietnam
- ______ 3. Watergate, resignation of President Nixon, succession of President Ford
- ______ 4. Women's Rights Movement (e.g., Betty Friedan, Gloria Steinem, and ERA)
- ______ 5. Energy crisis
- ______ 6. Human rights

Other Important Areas:

If there are other things you think are important for students to know, please list them here. Use the back of the sheet or attach additional pages if necessary.

(A blank page included for the reviewer's comments has been omitted here in the interest of cost.)

ATTACHMENT 13

AUSTIN INDEPENDENT SCHOOL DISTRICT
Department of Management Information
Office of Research and Evaluation

THE NAEP LEARNING OBJECTIVES

LITERATURE

______ I. Novels, Short Stories, and Plays
These words and their authors are significant because of universal characters, plots, and themes. This universality arises when a work portrays an experience or a characteristic common to humanity and so helps us to understand ourselves and to develop our personal values.

______ A. Novels
The emphasis is on American authors and their works (Twain's The Adventures of Huckleberry Finn, Cather's My Antonia). Major English and foreign (translated) works are also included (Robinson Crusoe, 1984, Don Quixote). Knowledge and understanding of characters, plots, and themes are essential.

______ B. Short stories
Most of the stories are by American authors (Edgar Allen Poe, O. Henry, Shirley Jackson). A knowledge of characters (Walter Mitty) and plots ("Rip Van Winkle") will be assessed.

______ C. Plays
The emphasis is equally divided between Shakespearean plays (Romeo and Juliet, Hamlet) and classical and modern plays (Oedipus Rex, Our Town, A Raisin in the Sun). Knowledge of characters, plots, and familiar passages will be assessed as well as an understanding of themes.

______ II. Myths, Epics, and Biblical Characters and Stories
Knowledge of these is necessary in order to understand other literature in our culture because of the frequency of allusions and references to these texts. This knowledge constitutes a cultural shorthand that enables students to recognize certain universal characters, symbols, and themes. With this knowledge, students may recognize the similarities of human experience, past and present.

______ A. Myths, heroes, and legends
Classical Greek and Roman mythology (Midas, Venus, Zeus) is the major emphasis in this area. Also included are the Arthurian legend (Merlin), fairy tales (Cinderella), folk heros (Robin Hood), and fables ("The Tortoise and the Hare").

______ B. Epics
Included are the earliest epics (The Odyssey) and those that came later (Paradise Lost).

______ C. Biblical characters and stories
Major biblical figures (Moses, Judas), events (the Flood), and parables (the Prodigal Son) are the core of this area.

______ III. Poetry
These poets and works are significant because of themes, allusions, and imagery that enable us to understand the meaning of our human experience, that help alter our sensibilities, and that enlarge our imagination. Some of these works are also frequently quoted and thus have become vehicles for communication with others.

Included are mostly American poets and their poems (Dickinson, Whitman, Hughes) as well as some English poets (Shelley, Blake). Some familiar passages (from Frost's "Stopping by Woods on a Snowy Evening") are also cited.

______ IV. Nonfiction
Some of these works and their authors are frequently referred to because they have come to embody the values of our culture. Others are worthy of study because of their success in the communication of ideas by the use of effective rhetoric.

A variety of American nonfiction has been included: speeches (King's "I have a dream"), historical documents (Declaration of Independence), and autobiographies (Thoreau's Walden).

Other Important Areas:

If there are other things you think are important for students to know, please list them here. Use the back of the sheet or attach additional pages if necessary.

(A blank page included for the reviewer's comments has been omitted here in the interest of cost.)

ATTACHMENT 14

ITEMS WITH MORE THAN FIVE ANSWER CHOICES

TO THE TEACHER:

Oops! The answer sheet we are using has spaces for only five answer choices, but some history questions have six answer choices. The test forms and the item numbers which have six answer choices are:

Test	Form	Item numbers
History	2	39-41
History	4	17-23

Please instruct students as follows:

ON FORM 2, ON ITEMS 39-41

Use the "A" bubbles next to items 69-71 as the "F" answer choice.

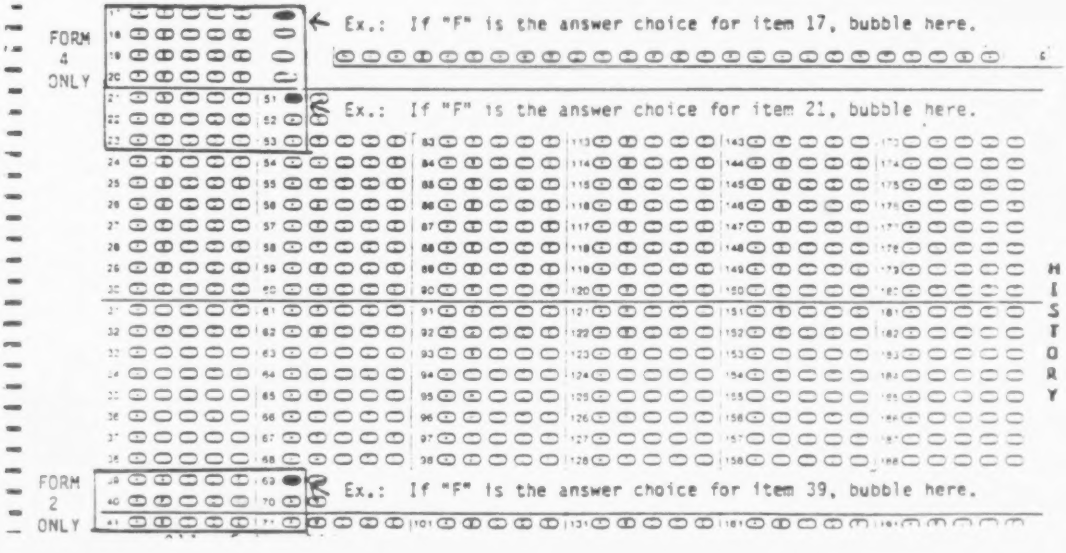
ON FORM 4, ON ITEMS 17-20

Mark an "F" answer choice in the blank space to the right of the line with the items. Line up the mark with the column of "A" bubble below.

ON FORM 4, ON ITEMS 21-23

Use the "A" bubbles next to items 51-53 as the "F" answer choice.

The examples below illustrate how students should mark the answer sheet. CUT SHEETS PROVIDED INTO HALVES AND GIVE EACH STUDENT A COPY.


FORM 4 ONLY
Ex.: If "F" is the answer choice for item 17, bubble here.
Ex.: If "F" is the answer choice for item 21, bubble here.
HISTORY
FORM 2 ONLY
Ex.: If "F" is the answer choice for item 39, bubble here.

All of the literature tests, and the other two history tests, have no items with more than five answer choices.

We are sorry for the inconvenience. An alert teacher caught this problem after testing began.

We are sorry for the inconvenience. An alert teacher caught this problem after testing began.

INDICATORS OF EQUITY AND EXCELLENCE: SEARCHING FOR CONSENSUS

Lee Hillman Brown
Jerry Litzenberger
Puget Sound Educational Consortium

INTRODUCTION

Fourteen separate public school districts in the general Seattle/Puget Sound area and the University of Washington form the Puget Sound Educational Consortium, established to promote mutual support among the districts and the infusion of university research into local district practice. Thus, from its inception, the PSEC has had the 1989 AERA theme as its primary function: promoting the interdependence of Research and Practice.

This is not a research report, nor does it present an evaluation of educational programs. In fact, the report stops short of any that I have heard presented at an AERA meeting. We, the Indicators Task Force, have spent an amazing two years talking about defining terms and gathering baseline data. This report is simply about gathering data: a document relating what we produced and how we did it.

A key ingredient of the Indicators Task Force was frustrating to some of us at first and turned out to be a great benefit: among the evaluators and student information people who conceived this project were placed -- by their superintendents, and to the evaluators' initial dismay! -- special program directors, curriculum directors, affirmative action officers, and even a person who had been the president of the district's teachers union. Because of the group's diversity, this paper has contributions from several different kinds of writers and undoubtedly changes style and tone from time to time. This attribute has been deliberately preserved in the finished product because the differences are important to the viability of the group and the group's product. Many of our communications to this audience are written in the form of individual essays. Taken all together, we hope we present to you part of the flavor along with the saga of our work.

Part of the strength of out product derives from the heterogeneity of districts as well as the heterogeneity of the district representatives. The districts range in population from 9,000 to 44,000 students and in ethnic minority population from seven per cent to 52 per cent of the total district population. Students eligible to receive free or reduced-price lunches make up ten to fifty per cent of the total enrollments. We serve communities with differing problems, from different administrative perspectives, and using different data systems. Because of the diversity, any products we can produce and find useful stands a good chance of being produced and used in other districts as well.

At this point, we are here to say that consensus is possible, but it is not necessarily elegant and certainly not easy! However, we believe that consensus of the type we have to present here is absolutely necessary if we are to be able to benefit from and believe each other's indicators. Through this AERA presentation, we hope to encourage others to undertake similar pursuits and we hope to establish contact with others who may be engaged in a similar kind of inquiry.

In this document, please find detailed descriptions of the project context, process, product, and recommendations for future work. In addition to the objective description of what was done and produced, the paper includes more subjective information for those wishing to benefit from our experiences. Among other topics, these include presenting information so that it can be understood, issues of disclosure, "catalytic validity", an "equity index", and a district view of the politics of collaboration.

MEMBERS OF THE INDICATORS TASK FORCE

DISTRICT	STAFF
Bellevue School District	* Jerry Litzenberger, Co-chair
Edmonds School District	* Lee Hillman Brown, Co-chair
Tacoma School District	* Linda Elmen
Federal Way School District	* Carole Davis
	* Kathy Purcell
Shoreline School District	* Jack Rogers
Everett School District	Ann Timm
Highline School District	Art Maser, Cathy Ross
Kent School District	Dian Blom
Lake Washington School District	Bob Lehman
Northshore School District	Lois Beal, Gary Reul
	Tom Scranton, Marion Weldon
Puyallup School District	Louise Moffitt, Steve Peretti
Renton School District	Gordon Pfleiter
Seattle Public Schools	George Sanders
Snohomish School District	Pete Dodson
University of Washington	Kathy Mueller, Ken Sirotnik

* Every one of these people contributed substantially to the indicators project. Those whose name is preceded by an asterisk contributed to the writing of the document.

PROJECT CONTEXT

From a suburban district's point of view, "disproportionality" traditionally has belonged to large city school districts. Certainly that was the case in the Seattle/Tacoma area until three years ago when one suburban district boldly disaggregated achievement information. Disaggregation -- reporting achievement data separately for the five major ethnic groups -- revealed that the different groups indeed show different achievement and further that nothing the district was doing during the public school years seemed to be working to narrow the differences between groups.

This event and the ensuing discussions lead to a decision by the area's superintendents involved in the Puget Sound Educational Consortium (PSEC) to establish a network of district staff to investigate Equity and Excellence. The project reported here is the group's major development to date: an information base in the form of individual district profiles that can be used to reveal equity (or the lack of) and to measure progress over the next decade in the direction of Equity and Excellence.

The basic pattern of profiling we follow extended on work that had been conducted by the Northwest Regional Education Laboratory (NWREL) in their "Onward to Excellence" program and in other school districts including Seattle, Philadelphia, Boston, and San Diego. The profiles are limited to student outcomes for this initial year. We intend for the process of sharing the information contained in these reports to be interactive. A sincere attempt will be made to identify the kind of information that communicates best with our educational audiences and to attempt to modify, to the extent that we can, our indicators and reporting formats for the 1990 school year.

Our intent in developing this reporting system is to determine the extent to which both equity and excellence exist in our educational programs. The philosophical premise we share in this information gathering activity is that program excellence cannot exist without equity. To the extent that the performance of the various groups we serve can be predicted based on the irrelevant variables such as gender, ethnic membership, or socio-economic status, we have not achieved equity. To support this perspective in producing a district profile, it is necessary to provide information on the performance of students from each of the different subgroups (ethnic, gender, socio-economic level). It is our hope that this information when presented to schools and the school community will present a unique view of the effectiveness of our programs and practices.

This represents merely the beginning of our efforts to provide good quality information that can be used as the basis for decision-making at both the school and district levels. By providing more information in a useable format to principals, teachers, and parents at the schools, we expect to help teachers make better decisions about how to serve all students. By compiling the information at the district level, we expect the school board and central offices to have a clear picture of progress being made in the district.

PROJECT PERSPECTIVES

As stated above, the project is being conducted from the following perspectives (which have been adopted as working assumptions):
(1) There are no systematic differences in human learning potential other than those attributable to individual variation. (2) Schooling environments can be created within which most students can develop highly their learning potentials with respect to a valued, common curriculum. In other words, ALL kids can learn.

With these working assumptions in mind:

EXCELLENCE is indicated by conditions, practices, and outcomes in schools that are associated with high levels of learning for students in all valued goal areas of the common curriculum.

EQUITY is indicated when there are no systematic differences in the assessments of these conditions, practices, and outcomes based upon race, ethnicity, sex, economic status, or any other irrelevant grouping characteristic.

WITHOUT EQUITY IN STUDENT OUTCOMES, THERE IS NO EXCELLENCE IN SCHOOLS.

These are the stated objectives:

1. To develop an information base (or comparable bases) for reporting equity and excellence problems and progress:
    a. To develop definitions of equity and excellence indicators,
    b. To determine how to obtain data from the different school districts,
    c. To present the data in a form which allows valid comparisons to be made between districts and which allows each district to examine the degree to which student outcomes indicate EXCELLENCE and/or EQUITY;

2. To continue the collection and publication of the indicator data for 5 to 10 years so the first year information forms the basis for evaluating progress toward the twin goals, equity and excellence; and finally,

3. To identify and disseminate programs which prove to promote equity and excellence.

The underlying, unstated primary objective -- the need we initially saw as most compelling -- was and is to increase awareness of problems of equity, especially in our suburban school districts. Before this project, all of our suburban districts felt equity was simply not a problem for them because they had no visible minority group and no one called attention to inequitable outcomes.

To date, we feel we have completed the first objective. In February 1989, eight of the fourteen member districts published the results to date of the effort to establish a common reporting format for data produced under common definitions. Two others are still working toward profile production. Two others have made conscious decisions not to participate, at least for this year.

Perhaps the most important outcome so far is the degree to which we have met our covert objective of increasing district awareness of equity issues. Teachers and building administrators -- even in districts which previously denied any equity problems -- are now asking for data on our indicators for their own buildings, and they are asking for programs that work. Now the real work begins!

METHODS

Because this work involves fourteen different school districts with fourteen different sets of problems and nearly as many information systems, plans and procedures need to be **practical**. We set about to extract measures from the school environment that already existed, not imposing any new measures. The indicators which emerged from a year's discussions is a flexible and diverse set of indicators that:

- teachers generally already keep and understand
- have the potential for use in the schools
- will be sensitive to change, showing change only when there is real change.

As of December 1988, the set included fourteen indicators. Twelve of these have been carefully defined (see Appendix); the operational definitions of #13, Discipline, and #14, State Student Questionnaire data, are still being refined. The indicators are:

1. Absence/attendance
2. Dropout information
3. Graduation information
4. High school course enrollment
5. Special class placement (tracking)
6. Achievement (statewide standardized test results)
7. Cumulative Grade Point Average (GPA) and grades
8. Washington Pre-college Test Information
9. Credit accumulation
10. Criterion-referenced test results
11. Communication (writing) skills
12. Student opinion about self and school
13. Discipline (expulsion/suspension)
14. State student questionnaire information

Frequent meetings, consensus building and revisiting, persistence, patience, and impatience comprise the basic techniques which proved to be most effective.

What seemed quite simple at first turned out to be a remarkably intricate problem. All we had to do was define a set of indicators so that each district can produce numbers which match the definition. We were very naive. We began with the easy one: attendance, important and easy to count. A concept as simple as "attendance" becomes amazingly complex. Some of the problems we discovered are typical of the problems with other indicators:

- Those buildings or districts which collect attendance information centrally actually tend to collect "absence" data.
- Most districts do not collect elementary absence information other than in the teachers' grade books so that central collection and disaggregation would be extremely costly.
- No two buildings could agree on a definition for a "day"! (At what point is a child absent for a day?)
- Did we mean all absences, unexcused absences, or absences unless a student was attending a school function such as a band trip?
- And, finally, the number we really need is the number of days absent divided by the number of days actually enrolled so that the absence rates are comparable for students enrolled 180 days and those enrolled for only thirty.

Each indicator involved such details and problems and the one year project slid into being a two year project. In the end, we settled on a practical definition for each indicator and one which each district could meet, a "lowest common denominator". This necessary concept frustrates each of us. We want to develop the TRUE definition rather than the lowest common denominator, but congruence even if inelegant turns out (1) to be necessary, (2) to be useful, and (3) to work. That is, both excellence and equity can be represented even with COMMON data.

A good approach to developing our indicators developed over about a year. We got so we would (first month) have a brainstorming session about in indicator, (between meetings) have one or two people work ~~meetings~~ to gather actual district data and write up a draft definition and presentation format, (second month) present the draft to the group and facilitate more discussion, (between meetings) then everyone would use the revised definition to gather their own district data. (Third meeting or month) This results in more ideas about how the definition must be refined and, possibly, consensus. At the rate of about one meeting a month, each indicator took about three months to develop. We only finished the twelve indicators because two or three indicators could be undergoing development at a time.

Key problems:

- Each team member had a managerial position in the home district. This was necessary in order to have the power to get the work done, but the busy district schedules often conflicted with the collaboration. The rest of the problems resulted from this one.
- Absenteeism -- resulting in no buy-in with the definition.
- Not doing homework -- no hands-on experience with the definition or in real problems preventing compliance.
- Different collection processes in the different districts.
- See the discussions below entitled Levels of Interaction, Politics of Collaboration, and District Support.

Helpful solutions (These may seem obvious now, but none was started soon enough in our work.)

- Develop notebooks with a section for each indicator.
- Assign primary responsibility for each indicator to one person.
- Tape record all discussion sessions or have a minutes taker who understands the concepts but who does not enter into the conversation (impossible, from our experience).
- Get detailed minutes of each meeting out almost immediately to all the team.
- Telephone everyone between meetings to ensure that assignments are understood and the next meeting is confirmed.

For any group of districts embarking on a project like this, please know that hard work, lots of time, and periodic frustration and touchy feelings seem to be unavoidable parts of the process. We have tried to portray some of the issues and our recommendations in the pages which follow.

THE POLITICS OF COLLABORATION

The Equity and Excellence task force is composed of representatives from the consortium districts. The representatives were both program and research people, which makes for an interesting mix. These people brought to the task force diverse backgrounds and experiences.

While program, research, and evaluation specialists attempted to reach some common ground they each were having to go back to their respective districts to encourage support and reduce fears associated with this data gathering activity. Equity and excellence are sensitive issues for school districts to deal with. No district wants to be known as having a program that is not equitable nor excellent. Any hint of such raises the anxiety level.

Therefore superintendents, principals, and others were concerned about the use of the resulting data. How the data would reflect on the districts was a concern expressed by people in each of the consortium districts. Needless to say the whole issue became a political one.

The task force not only had to complete their stated charge of identifying indicators of equity and excellence but became responsible for development of politically sensitive strategies to encourage the districts to move forward. With the stated and unstated charge, the task force members began the process of collaboration.

Collaboration is normally a time-consuming process. When the individual members of the group have differing points of view and represent districts each having a unique set of characteristics, the process of collaboration became more difficult. During the many months, the task force members worked their way through phases of disagreement, heated discussion, excitement, and finally agreement and consensus. The unity of the task force members has produced some unexpected results:

- Districts with greater computer capabilities assisted those with lesser capabilities.
- Program people and research people learned to talk the same language.
- Three districts created positions to handle the data gathering activities.
- The strong feeling of oneness of purpose developed within the group.
- Departments within districts have developed joint efforts toward a structured data system.
- Universities and other agencies have volunteered their assistance.

On the down side, several task force members still had their job responsibilities and had to do much of their task force work on their own time.

Due to the deep commitment to equity and excellence, the first year objectives were met. The districts completing the data gathering for this first effort are all taking a positive proactive stance. They are saying we are looking at our district data as baseline in planning programs which will have a greater effect on students. Positive student outcomes are the major purpose of the focus on equity and excellence. Even though each district is at a different place, the task force is determined to continue its work while assisting districts in a collaborative manner to see that students are provided an education where equity and excellence are the top priority.

DISTRICT SUPPORT

Getting to this simple list of indicators and publishing simple district profiles TOOK AN AMAZING AMOUNT OF WORK. The group began as a mixed group of volunteers and draftees with differing degrees of district support and one all-encompassing similarity: naivete.

We had

- the naive assumption that once committed, our superintendents would stay committed;
- the naive assumption that we would be given time for this work; and
- the naive assumption that there is a correct definition for simple constructs such as "absence"!

District support for this initial year in gathering data for the PSEC Indicators of Equity and Excellence was mixed. Initially the executive board of the PSEC consisting of the fourteen district superintendents and the University of Washington representatives voted unanimously to support the action plan proposed. However, by the time the project was underway 8 of the 14 superintendents were new to their districts. This certainly led to the need for each school district to review its commitment to the project. In general, district support ranged from enthusiastic, unquestionable commitment to outright hostility. Most districts found themselves somewhere in the middle.

Even when superintendents were committed to the success of the project, the commitment came in different forms. Commitment of staff time to the project was often a problem. In one district, four staff members were assigned to various parts of the project. Two of these staff were given extra days to carry out their part of the project, and the team was able to hire a clerk for three months to compile information for the profile. In other districts part-time staff were hired to carry out various aspects of data collection, analysis, or presentation. However, in the case of most of the districts, those responsible were expected to carry out their part of the project along with the rest of their responsibilities. In several districts there was only one person assigned to the project. Most of the participants complained about wearing too many hats and not having enough time to devote to the gathering, analysis, and presentation of the data. At least one participant was never clear on the status of the project in his district. Without direct access to the superintendent, the participant managed to compile those data available and prepare a summary report, never being clear about whether this was seen as a priority use of district time.

A second major area of district concern was in the availability of data. In a few cases, all of the data were available through the district computer system and relatively "simple" to gather in disaggregated form. In other districts, the process was much more complex and required either the hand gathering of data or, in one case, the creation of a special data tape which would allow much of the data analysis to be done. In some cases the data might be available to be disaggregated by gender or by ethnicity, but not by gender within ethnicity. Only two districts were able to disaggregate by an indicator of socio-economic status (eligibility for free or reduced-price lunches). In several districts the data simply were not there. One district representative commented that he not only lacked a computer of his own and time to get the work done, but that many of the data he wanted had simply never been entered into the computer. Even when data were available, such as MAT testing in grades other than those tested by state mandate, this representative had no means of matching test scores to other demographic information since no identification numbers were used in the testing.

LEVELS OF INTERACTION

Although members of the indicators task force represented their district, their participation within the group was governed by the highest of professional interest in a project that would ultimately benefit all students. Representatives, therefore, formed a unique group that held rich discussions about the qualities and characteristics of indicators of excellence and equity:

- What is the purpose of the indicator?
- How useful will the information be?
- Who will use it?
- Can/will it be used wisely?
- What is the best way to describe the indicator? to table the data? to illustrate/graph the data?
- What is the best source for the data?
- What is a common source for the data?
- Where is the information located in the home district? Who has it? In what form?
- How close can the district report be to the boiler plate?
- If the data are too controversial, what will happen?
- Should the indicator reflect state, national or other "expert" definitions?

As indicator definitions were drafted, agreeing on a definition was complicated by review and re-evaluation. Minutes, notes, and group persuasion were used to keep everyone abreast of the most recently determined best definition.

The universe established through the consensus-building activity of the task force is but one universe representatives found themselves working in. Most are facile with the universe that is their district. However, district idiosyncracies affect task force productivity. For example, the priority of the indicator project varied significantly among the districts. Some information is readily available in some districts, more elusive or not available in others. Some information sources were technically more reliable than others: some districts are computer-supported and data dependent; others have little or no data management systems. Most districts' capability of manipulating data to meet task force standards was limited.

A parallel universe to that of the districts is the Puget Sound Educational Consortium (PSEC) itself, a complex collaboration among school districts governed by the superintendents. This was a consortium project. Were we doing what they really expected?

Beyond the universes of task force, district, and consortium, are state and national forces where the universe includes government regulation and the mass media. Task force members certainly considered their understandings of this larger context as they developed definitions.

Perhaps some generalizations are now in order. First, in each universe it was necessary for task force members to continue an education process about the selected indicators. The examination of definitions, the availability of data, and the significance of the study were both hotly contested and seemingly ignored depending upon the audience. That is, the ideas continued to be negotiated within each universe.

Second, ideas were also continually exchanged among universes. For instance, district positions on indicator matters affected the work of the task force; similarly, consortium governance (superintendents' meetings) affected timelines. Third, the task force was frustrated frequently by settling for the "lowest common denominator". Sometimes the lowest common denominator was the data source (such as a state report), or the manipulation (information not available by gender within ethnicity), thus limiting the consensual definition. The process was thus dynamic, slower than projected, and fraught with ambiguity.

LIMITATIONS AND APPROPRIATE USES FOR DISTRICT PROFILES

Our reports present descriptive information about our districts and schools. The information we have collected is not intended to be an evaluation of the district programs, practices, or individual schools within the district. Research designs and analyses of the information collected based on specific program objectives would be necessary before this information could be used to evaluate programs. The proper use of this information in the long term will be to use information collected in this initial year as a baseline against which future changes in individual schools and programs can be monitored.

In a monograph entitled "Educational Indicators: A Guide for Policy Makers", Jeanie Oakes (1986) describes proper uses of indicators. These include the following:

- Comparisons are inappropriate unless the indicators are defined and measured in the same way.
- Indicators can track schooling trends if they measure enduring features of the system and are consistently defined. However, indicators will not, by themselves, tell us what caused the trend.
- When research has clearly linked indicators and other performances in the educational system, they can predict future performance quite accurately. More often, indicators may reveal patterns of change that allow us to predict with somewhat less certainty.
- Indicators will bring considerable new knowledge to bear on educational issues, stimulate debate and suggest new approaches, but they will not provide unequivocal judgement about the quality of schooling or direct answers about how education should be improved.

The information contained in our profiles represents one set of indicators of equity and excellence. For each perspective presented, at least ten others were discussed and rejected, and probably that many more are possible. Over time we will refine these perspectives based on feedback from our several audiences. The inclusion of thoughtful feedback and refinement will result in valid indicators of equity and excellence.

SHORT AND LONG TERM RESPONSES TO DISTRICT PROFILE USES. We have attempted to avoid drawing conclusions about programs from our profiles. We are encouraging each district (and each school in the districts which produced school level information) to come to its own conclusions about the significance of the information contained in their profile. The important issue is that schools begin to learn how to use objective information to make important educational decisions. It is likely during our first year that people will go through a period of discomfort as they become familiar with the data. We feel this is an important step in preparing for productive problem-solving and will lead to a properly balanced perspective on a use of profile information.

In the long term, we are confident that there is enough evidence based on existing research to show that schools can become institutions that respond to all students' learning needs. However, we also believe that this will not happen if schools continue to operate and be structured as they have been in the past. Only if we make major changes will we address the issues of equity and excellence. Multi-year trends will be our best source of information about the success of our change efforts. Our initial reports need to be seen as one piece of information in a larger puzzle, not as the source of immediate answers. To the extent profile information is considered in context and over a sufficient time period, we are confident they will help our districts reach our goal of an excellent education for all.

RESULTS / CONCLUSIONS / POINT OF VIEW

This project has had a single major result with three manifestations. The major result for folks outside our consortium is the set of indicators, carefully defined, and the agreed-upon presentation formats for each of them. From each participating district's point of view, the project's major result is the individual district profile recently published. Finally, for the superintendents as a group and for the newspaper journalists, the major result is the synthesis report, also recently presented. The hefty appendix to this document details each of the indicators and provides the synthesis summary of the data presented by all the participating districts.

ACCEPTANCE AND USABILITY OF THE DATA. The representatives to the Indicators Task Force experienced a growing concern from curriculum leaders in their own districts about the nature of the equity and excellence studies. Several administrators expressed dismay that the report dwelled on strictly negative statistics, pointing out inequities in student performance while doing nothing to indicate positive outcomes for the majority of students. Further, the criticism was noted that the indicators may reflect conditions over which the schools have little or no control, and that to break out subgroup statistics only promotes unfair criticism of the schools.

On the other hand, each district has experienced the interested reaction and curiosity of a select few that is necessary for the topic of equity to gain stature. We sense a sort of "catalytic validity" to our work: it is valid because it instigates conversations that need to take place. The topic seems to have a power of its own. In fact, the evaluators in the group have all remarked that this information is picked up and looked at by administrators who have refused to look at any kind of data all these long years! Finally, most districts have not yet produced building-level profiles partly because of time and partly because of interest. Yet in every district, when shown the district profile, the staff of some buildings request similar outcome information for their own students. These have moved from annoyed indifference to genuine interest.

AGREED FORMATS. Each indicator is presented in the profiles in a specific format designed so that each kind of reader would find information presented in their preferred style. The format includes a rationale for the indicator; definition including data sources and formulae; data weaknesses; then the data in number form, graphic form, and in summarizing statements. The first three sections of each indicator are provided in detail in the appendix. The section pertaining to cumulative grade point average is complete with actual district data to illustrate a typical finished chapter.

DISTRICT COMPARISONS. Overall, it is apparent that there are both differences and similarities among districts. For example, districts differ considerably in reported attendance rates from 2.7 days absent in one district to 12.7 days absent in another; enrollment percentages in advanced math courses ranged from 40% in one district to 10% in another; and so forth. Variation between districts is likely due to interactions of a complex set variables including curriculum and instruction differences, district demography, policies, and record-keeping accuracy. Moreover, not all indicators were reported by all participating districts. For example, attendance data were available for only five districts; some districts had overall attendance data but were unable to disaggregate it.

In spite of these district differences, there are pervasive patterns within districts once the data are disaggregated by ethnic group, gender group, or socio-economic status. Furthermore, simple indicators show the same equity patterns as complex ones when summary data are disaggregated.

Regardless of the indicator or the district, the data collected for Black, Hispanic, and Native American students typically reflect less favorable results (higher absence rates, lower GPAs, higher enrollments in remedial programs, lower enrollments in advanced courses, higher dropout rates, lower test scores, and so forth) than Asian and White students. The overall magnitudes of averages, percentages, and statistical differences vary somewhat from district to district but the patterns remain the same.

Consistent patterns for females and males are also apparent, but are, with few exceptions, less dramatic in frequency and magnitude. Nonetheless, the trends are predictable: female students tend to be absent more often and are enrolled less frequently in advanced math and science courses. Male students, however, tend to have lower GPAs, higher dropout rates, lower enrollments in advanced English courses, lower graduation rates, lower scores on language portions of standardized tests, and remarkably higher enrollments in special needs programs.

Finally, there are marked differences in performance on the indicators in the two districts that were able to disaggregate information for economic status groups. Students eligible to receive free or reduced-price lunches have higher absence and discipline rates, higher percentages enrolled in remedial programs, lower percentages graduating from high school, and lower test scores and grade point averages than other students.

In sum, there are systematic differences in student performance based on ethnicity, gender, and economic status within and across districts on each indicator we used. In accord with the above definitions and working assumptions, these differences clearly indicate inequity in the educational experiences for students. If equity is required for excellence, then our schools are not excellent. Furthermore, inequity is not just an urban phenomenon.

We are heartened by the publication of the profiles and the end of this phase of the project. We are disheartened with the specific outcomes. We have found no district in which there is equity. And now there is more work to be done.

OTHER OUTCOMES OF THE PROCESS.

We have learned several other lessons from the process itself. Because we seek to encourage other groups to join in this type of adventure, a few of these are listed here.

PUBLIC DISCLOSURE laws are fact. As soon as data are tabulated, the community has a right to know about them. We have learned that assertive pursuit of media coverage and openness to disclose our findings to the public creates a beneficial atmosphere. Journalists will get the data eventually, and if we give it out freely, we win by enjoying better press. This stance, by the way, runs counter to the advised operation of most district Public Relations offices. It may be the most controversial stance to take in many districts. However, it happens to work.

DATA REQUIRED BY OUTSIDE (FUNDING) AGENCIES provide us with ready-made sources of information which is more likely to be comparable than most. Eight of our ten complete indicators derive from this type of requirement.

CONDITIONS OF COLLABORATION. The single most important precondition to the ability of our fourteen school districts to work together to produce this consensus work is the umbrella of the Puget Sound Educational Consortium. The results to date represent, to some extent, a test of John Goodlad's recommendation that schools and colleges should develop networks around issues of common concern. Goodlad and the University of Washington School of Education provided the atmosphere necessary for the Consortium to develop.

SENSE OF EMPOWERMENT. The successful collaboration has created in the Indicators Task Force a real sense of expertise and strength in the state educational community. Perhaps because we represent over a third of the state's students or perhaps because we have grappled with difficult questions, we feel we have been empowered.

NEXT STEPS

The indicators reports are not intended to be definitive or complete this first year. Several indicators are not yet feasible to collect for all districts. For example, districts must develop systems for collecting longitudinal or cohort dropout and graduation rates before these statistics can become indicators. Cross-correlations of such factors as family structure and socio-economic status need to be incorporated in future annual updates. It is anticipated that the initial study of these data will serve to promote further study and sophistication in data gathering. As districts identify and address equity issues, more and more questions will arise, requiring additional information and data sources. The outcome should be greater knowledge and understanding of our student clientele and, ultimately, appropriate instructional measures taken to provide equity in student outcomes.

Technical work must include at least the following:

- Review each indicator, remove inconsistencies in reporting and keep or decide to eliminate each one.
- Develop the information and record-keeping capabilities in each district.
- Develop building-level reports and get the information published so it is meaningful where the educational action is.
- Maintain records and develop longitudinal reporting.
- Consider the addition of indicators such as grade retention.

Work toward Equity and Excellence in our schools is the real purpose of all this activity. Some districts and schools are already experimenting with innovative activities and programs targeted toward more equitable and excellent education. As a direct result of the consortium efforts, two major conferences have been held in this immediate area which have focussed on equity curricula. We need to make concerted efforts to seek out all that we know about these innovations, get some going in our buildings, and document what occurs. We do not want to end up five years from now with a well-maintained, longitudinal database but no information about programs and practices that work.

INTRODUCTION TO THE INDICATORS

District profiles include the title page for each indicator that was developed through the consensus process. Each title page includes (as on the following pages) three sections entitled Rationale, Definition/Data Sources/ Formulas, and Data Weaknesses. The Data Sources section might include actual sources for the particular district, but the definition and formulas should all be identical (emphasize should).

Following the title page for each indicator, the district profiles each include a table of numbers showing the district results for that indicator and then a graphic representation of those results.

Finally, most districts completed the section by adding three or four verbal descriptions of the findings, preferable in an objective rather than valuative tone.

Examples are included here from actual district profiles to clarify. These examples have been chosen specifically to illustrate the three general ways developed for looking at equity in the districts. These were to:

1) Compare the percentages of each ethnic group which met the criterion,
2) Compare subgroup's percentage of the program to the subgroup's percentage of the whole population, or
3) Compare medians or means on the indicator for each subgroup.

Each method works for some of the indicators. Appropriateness of each method of presentation seems to become clear only after analysis people actually try to put the indicator information into each of the forms.

To illustrate the actual kind of data presentations our profiles include, two indicators are accompanied by sample data tables and graphics from actual district profiles. These are course enrollment and cumulative grade point average.

ATTENDANCE

Rationale

School attendance is an important indicator of program quality. Used in conjunction with other indicators such as achievement and special class placement, trend data on absence can help to predict at-risk groups who are potential dropouts. The use of this information is more valid if definitions at the school level are consistent over years. To the extent the information is collected in comparable ways over time, this indicator can serve as a measure of program excellence. Decreasing absence rates can certainly lead to more instructional time and should result in higher performance measures on other indicators.

Definition/Data Sources/Formulas

Our definition of absence is the number of total student absences divided by the total number of students enrolled in the relevant grades. Elementary, middle school, and high school absences should be compiled separately.

An alternative definition (improved) would take into account the number of days each student is enrollment during a year. Because of the difficulty in maintaining the number of days absent for the number of days enrolled, this temporary definition was selected. In future years, with more sophisticated computer assistance, we may be able to produce more accurate indications of school absence.

Data Weaknesses

This information is collected using different procedures in each school and each district. The reliability of the indicator is lessened to the extent different policies are followed at each school. The definition of an absence particularly at the secondary level where class level absences are maintained is easily misinterpreted. Elementary school data is particularly difficult to maintain unless attendance is captured from grade reporting information.

At the high school level, students may be (are) enrolled for different numbers of classes each day. The most accurate calculation for this level may be one taken from actual transcript information, counting absence in each class and dividing by the total number of student classes. In this calculation, James Jones might take six classes and so would be counted six times for deriving the schoolwide or districtwide average. Abigail Johnson might be enrolled in four classes. A "day" or a half day would be very different for these two students. However, separate class absences divided by total classes (not kids) would probably yield the best absence information, "best" in the sense that the information might best represent actual attendance patterns. But the consortium districts have not yet explored the practicality of this approach.

Only one district reported elementary absence data or middle school absence data.

Findings of the 1988/89 School Year

High School absence data were available for five of the reporting districts. The average days absent for total high school populations varied dramatically from district to district. In the district with the lowest average, students were absent, on average 2.7 days per year. In contrast, in the district with the highest average, students were absent 12.7 days per year.

- In four out of five districts, Asian students were absent least.
- In three out of five districts, Native Americans were absent most.
- In general, females were absent more than males.

DROPOUT INFORMATION

Rationale

Dropout statistics provide a clear measure of school excellence. Any decrease in dropout rates that is real, and not just an artifact of changing the way we count dropouts, provides a measure of the increase in school excellence. Equity in dropout rates would show that the schools are not better for some groups of students than for others. Lack of equity (which is the case nation-wide) is indicated by higher dropout rates for some groups than for others. Identifying the groups with higher dropout rates may help focus dropout prevention efforts.

Definition/Data Sources/Formula

Books have been written on defining dropouts and on calculating dropout figures. We recognize that the definition and calculations presented here are potentially controversial, especially because they give "cross-sectional" information about dropouts rather than an analysis of dropout values for each graduating class. However, (l) The data are produced each year for the Washington State Superintendent of Public Instruction (SPI) and can be compared to those from other districts in the state; and (2) Equity and Excellence can both be addressed using the data if the definition is kept constant over the years. In other words, we decided to stop arguing definitions and produce data that are accurate and consistent in their own right, even if not absolutely perfect.

For this report, all students counted in the district's enrollment (Special Education and ESL students included) are counted as dropouts if they leave school and do not return by the following school year for any of the following reasons:

School Related

_Expelled or suspended without returning or transferring to another school
_Had poor grades.
_School was not for me.
_Completed grade 12 without completing graduation requirements.
_Did not get into desired programs.
_Could not get along with teachers.

Family-related

_Married or planning to marry.
_Pregnant or had a baby.
_Had to support family.
_Chose to stay home.

Peer-related

_Friends were dropping out.
_Could not get along with other students.

Other

_Offered training/chose to work
_Wanted to enter the military
_Moved too far from school
_Wanted to travel
_Other or unknown

Cross-sectional dropout rate:

$$\frac{\text{(\# students in grades 9 through 12 dropping out between Oct. l and Sept. 30)}}{\text{(Total \# of students in grade 9 through 12 on Oct. l of the first year)}}$$

This is the information which we report to the state office (SPI) in the fall. (# = grade 9 students + grade 10 students + grade 11 students + grade 12 students). The formula OVER-STATES the real ONE YEAR dropout rate because the numerator is cumulative, counting ALL dropouts during a school year even if they transfer into the district during the year, but the denominator reflects enrollment at one date, not cumulative enrollment. On the other hand, the formula UNDER-STATES the four-year percentage of students who drop out of a given high school class over four years.

Accuracy in recording reasons for leaving school is essential to valid dropout statistics and the probability of 100 percent accuracy is low. However, to maximize accuracy, it is important to have each building in the district review the list of students who have been counted as district dropouts.

Data Weaknesses

- We have no way of tracking the high school outcomes of the students who transfer to other schools.
- Dropout figures include students who say they are moving to another school district but whose academic records are never requested by the receiving school. Some of these presumed dropouts may indeed have continued their high school education elsewhere.
- Data accuracy depends on the accuracy of the computer files which depend on (l) the building's success in tracking down students who leave without announcing intention to leave; (2) the building's zealousness in updating the computer files with any new information
- This reporting time-frame forgets the students who drop out before entering high school at all. These kids leave the system before they are even counted.
- The formula underestimates the extent of the dropout phenomenon as it impacts a given high school class over four years.
- The true dropout picture emerges only when students are followed for several years, at least from the year before the transition to high school.
- Students who complete four years but do not receive diplomas are rarely called dropouts. Records would have to be continued through the fifth high school year to determine which students continue and which drop out.

Findings of the 1988/89 School Year

Eight school districts reported dropout data. The annual cross-sectional dropout rate is deceptive. Districts indicating a five percent dropout rate annually actually may represent over twenty percent for a given graduating class in a four year high school period.

- Black, Hispanic and Native American students, in most districts represented, tend to have a disproportionally higher dropout rate than either White or Asian students.
- Asian students tend to have a disproportionally lower dropout rate than Whites, other minorities and the district's average.
- Native American students tend to have the highest dropout rate of any ethnic minority.
- Male students tend to have a slightly higher dropout rate than female students.

GRADUATION INFORMATION

Rationale

Graduation is an overall goal for our instructional programs. Therefore, it is valuable to look at the percentage of each class which receives a diploma on time. If programs are excellent, we ought to see a high percentage of students graduating. If programs are equitable, there ought to be no systematic differences in the graduation rates for students based on ethnicity, gender or socio-economic status.

Data Sources/Definitions/Formulas

Once again, the state-required definition has been adopted for this indicator. SPI requires that each district report the number of graduates by gender and ethnic group. The state report data form the basis for this chapter.

Formula:

$$\frac{\text{number who receive diplomas in each sub group}}{\text{total number who graduate}} \text{ X } 100$$

This gives the graduate percentage for each sub group. For determining equity, this percentage must be compared to the subgroup's percentage of the total high school population (grades nine through twelve).

Example (hypothetical). Say that 40 Asians graduated in a total graduating class of 100. According to the calculation presented here, indicator percentage is 40% for Asians. 40% of the graduating students were Asian. This by itself tells us nothing about equity. To determine equity, we need to compare this to the percentage of the total high school population is Asian. To complete the example, in the total high school population (grades 9 - 12) 100 of 500 students are Asian, or 20% of the total school is Asian. Now we can tell that Asians are over-represented in the graduating group by a factor of 2.

The need for this rather complicated process for evaluating graduation information becomes clear as we examine ethnic group patterns in leaving school. We find that both mobility and dropping out of school are related to ethnic subgroup. In order to produce the most clear graduation statistic, the group percent (or in this case, percentage of all senior Asians who graduate) would present a biased picture made invalid by fewer Asians than other minority students dropping out before the senior year. In districts such as the suburban districts in the Consortium, minority student group numbers are small and therefore likely to vary considerable from year to year, and one early dropout will make a difference.

Data Weaknesses

The true picture of graduation rate will only emerge as we are able to identify the progress of each class throughout the high school years. What proportion of the freshman class stays to graduate?

Findings of the 1988/89 School Year

Seven districts reported graduation information. Ethnic percentages of graduating classes has been compared to ethnic percentages of the total high school population. Graduation statistics parallel dropout rate profiles, but not to the same degree. Asian students are slightly over-represented among graduates while significantly under- represented among dropouts. Blacks, Hispanics and Native Americans are under- represented among the graduating class and over-represented among dropouts. Males tend to be under-represented in the graduating class.

COURSE ENROLLMENT

Rationale

The issue of course enrollment or access emerged from gender equity research. Enrollment patterns in upper level math and science courses have consistently shown that women and students in at least some minority groups are not enrolled in the expected proportions. Scores from college entrance testing have been lower for these two groups probably due to the smaller number of students who enroll in the courses that best prepare students for our colleges and universities. These scores are best predicted from courses taken in high school and gender and ethnic differences on college entrance tests undoubtedly result from course selection more than any other single factor. By monitoring the course enrollment patterns in our schools, we can determine whether there is equity in course access and if equity does not exist, this information can serve as feedback to help us erase any artificial barriers to enrollment in our upper level courses that may exist.

For this review, a sample of upper level courses was selected from math (third year and above), English (fourth year), foreign language (third year and above) and science (chemistry and physics). Our sample was taken from all students enrolled in fall quarter.

Definitions/Data Sources/Formulas

To determine enrollment equity, take the percentage of all enrolled students who belong to a given subgroup and compare this to the percentage of special class enrollment which belongs to that subgroup.

$$\frac{\text{\# Subgroup students enrolled in selected classes}}{\text{Total \# Subgroup students enrolled}} \times 100$$

This value can be compared to the percentage each subgroup represents of the school and district population. This comparison will reveal whether proportionality exists in access to math, science and English or foreign language courses selected. For districts to value the information, it is necessary to identify the courses that should be included with input from teachers and administrators in each district. This necessarily results in some differences between districts, but "college bound students probably do take somewhat different courses in different districts. Furthermore, having the districts determine courses works, as we see in our outcomes.

Weaknesses

This method of determining equity of course access will reveal current proportional or disproportional services to the several groups of students we serve. What this information will not provide is a specific identification of what the access barriers may be. A much more detailed review of advising students on course enrollment as well as tracking procedures within school will be necessary before an effective change process can be developed.

Findings of the 1988/89 School Year

Math
Data for enrollment in upper division high school math courses was available for all seven districts. The definitions varied somewhat from courses including Algebra 3 and above to also including honors classes in Geometry. Although the magnitude of the percentage of students enrolled in the various districts varied, from a high in one district of 40 percent of high school students overall to a low in another of approximately 10 percent, the patterns within ethnic and gender groups were generally similar. In almost all cases a larger percent of Asian students were enrolled in these math classes than any other ethnic group. In three districts, Black students had the lowest percent of enrollment, in three it was Native Americans. In one district, both Blacks and Native Americans had the lowest percent of students enrolled. In five out of seven districts there were a slightly higher percentage of males enrolled than females.

Science

Data for science enrollment varied from district to district. In one district only physics was included. In another, courses included advanced ninth grade biology, any honors courses, physics, and any second year or advanced topic science course. Once again, however, the same patterns repeat across districts. In still other districts, courses included chemistry, physics and any advanced topic or second year science course. In six out of seven districts, a higher percentage of Asian students were enrolled in advanced science classes than any other ethnic group. Hispanic students had the second highest percentage of students enrolled. In four districts, Native Americans had the lowest percentage of students enrolled, and in one district the lowest percentage was for Black students. In one district, no Black or Hispanic students were enrolled in any advanced science class. In five out of seven districts, there were a slightly higher percentage of males enrolled than females.

Foreign Language

Data on enrollment in foreign language classes for students in third year and beyond was reported for six districts. Participation in higher level foreign language ranged from 16 percent of total high school enrollment in one school district to 4 percent in another. In three districts, Asians had the highest percentage of enrollment, while in three others Hispanics had the highest percentage. In three districts no Native Americans were enrolled in advanced level foreign language classes. Females tended to be more often enrolled in third and fourth year foreign language classes than males.

English

Definitions of English classes varied from district to district. Data was available for six districts. Nonetheless, similar patterns emerge. In four cases Asians had the highest percent enrolled. In one case Native Americans had the highest percent enrolled, in yet another it was Hispanic students who had the highest percent enrolled. In two cases, no Native American students were enrolled in advanced English classes. In five cases Native Americans had the lowest percent of enrollment in these English classes. In general, females were slightly more likely to be enrolled in advanced English classes than males.

NOTE: Following please find a sample of one district's data table and graphic relating to this indicator. These show an example comparing subgroup enrollment in each area to district enrollment figures.

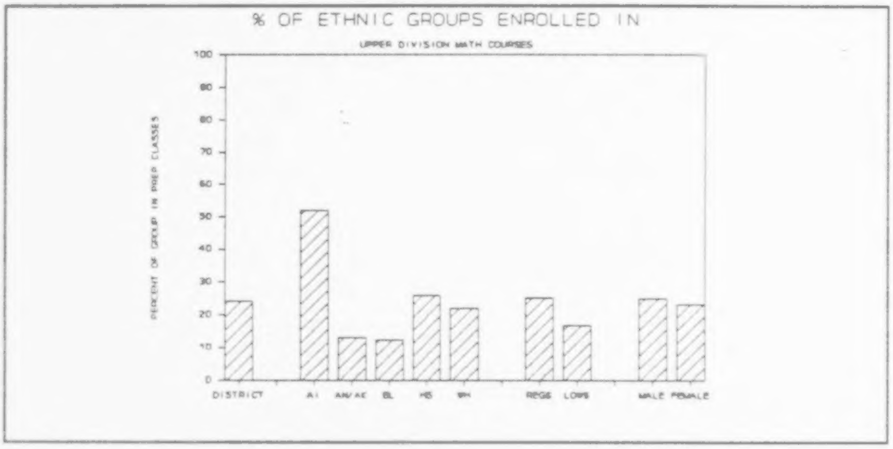
"UPPER DIVISION" OR COLLEGE BOUND CLASSES

Compare subgroup enrollment in each of the subject areas to these district enrollment figures.

	DISTRICT	AI	AN/AK	BL	HS	WH	REG$	LOW$	MALE	FEMALE
HIGH SCHOOL ENROLLMENT	5242	417	84	81	77	4583	4534	718	2682	2560
Percentage of Total		8.0%	1.6%	1.5%	1.5%	87.4%	86.5%	13.7%	51.2%	48.8%

"COLLEGE BOUND" MATH CLASSES

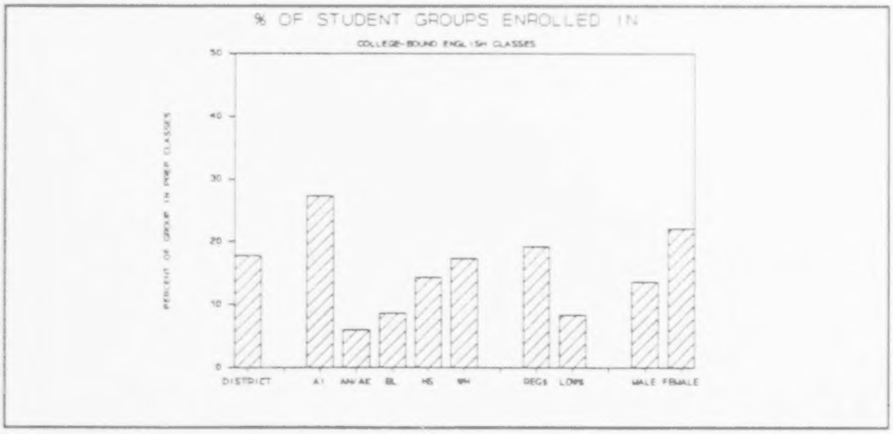
INCLUDES CALCULUS, SECOND YEAR ALGEBRA, PRE-CALCULUS, TRIGONOMETRY, MATH ANALYSIS, AND HONORS GEOMETRY.



	ALL	AI	AN/AK	BL	HS	WH	REG$	LOW$	MALE	FEMALE
PREP MATH ENROLLMENT	1270	217	11	10	20	1012	1149	121	674	596
PROGRAM %		17.1	0.9	0.8	1.6	79.7	90.5	9.5	53.2	46.8
GROUP %	24.2	52.0	13.1	12.3	26.0	22.1	25.3	16.9	25.1	23.3
EQUITY INDEX:		2.1	0.5	0.5	1.1	0.9	1.0	0.7	1.1	0.9

"COLLEGE BOUND" ENGLISH CLASSES

INCLUDES:
"COLLEGE PREP" ENGLISH CLASSES AT ALL GRADES, PLUS CP COMPOSITION, ADVANCED PLACEMENT ENGLISH (GR 11 & 12), CREATIVE WRITING, AND SENIOR HUMANITIES.



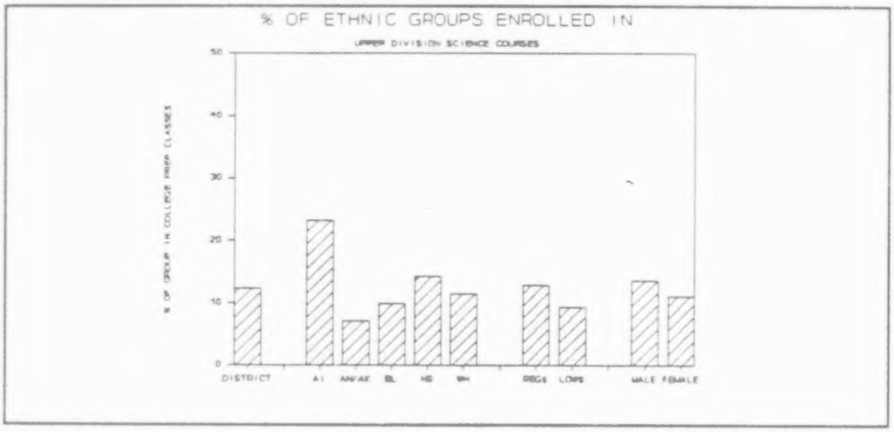
	ALL	AI	AN/AK	BL	HS	WH	REG$	LOW$	MALE	FEMALE
PREP ENGLISH ENROLLMENT	930	114	5	7	11	793	870	60	365	565
PROGRAM %		12.3	0.5	0.8	1.2	85.3	93.5	6.5	39.2	60.8
GROUP %	17.7	27.3	6.0	8.6	14.3	17.3	19.2	8.4	13.6	22.1
EQUITY INDEX:		1.5	0.3	0.5	0.8	1.0	1.1	0.5	0.8	1.2

Compare subgroup enrollment in each of the subject areas to these district enrollment figures.

	DISTRICT	AI	AN/AK	BL	HS	WH	REGS	LOWS	MALE	FEMALE
HIGH SCHOOL ENROLLMENT	5242	417	84	81	77	4583	4534	718	2682	2560
Percentage of Total		8.0%	1.6%	1.5%	1.5%	87.4%	86.5%	13.7%	51.2%	48.8%

"COLLEGE PREP" SCIENCE CLASSES

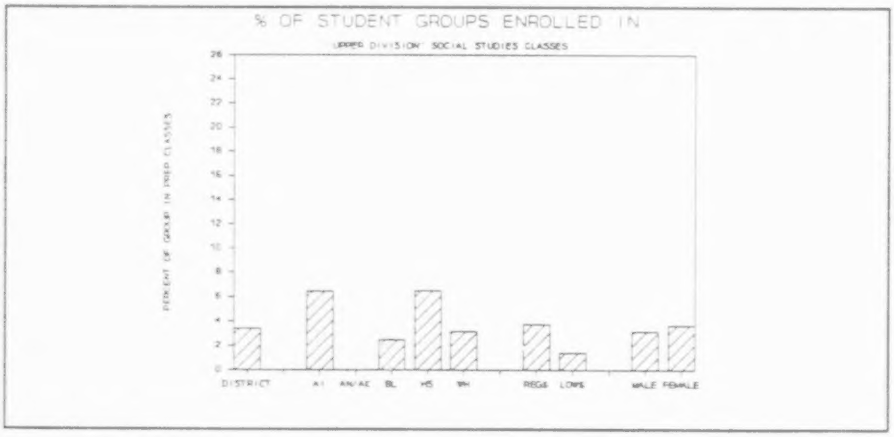
INCLUDES:
BIOLOGY: NINTH GRADE, SECOND YEAR, AND "HONORS";
CHEMISTRY: HONORS & SECOND YEAR; PLUS ADVANCED CLASSES IN ENVIRONMENT, PHYSIOLOGY, ZOOLOGY, INVERTEBRATES, VERTEBRATES, PHYSICS, AND LAB ASSISTANTS.



	ALL	AI	AN/AK	BL	HS	WH	REGS	LOWS	MALE	FEMALE
PREP SCIENCE ENROLLMENT	650	97	6	8	11	528	583	67	366	284
PROGRAM %		14.9	0.9	1.2	1.7	81.2	89.7	10.3	56.3	43.7
GROUP %	12.4	23.3	7.1	9.9	14.3	11.5	12.9	9.3	13.6	11.1
EQUITY INDEX:		1.9	0.6	0.8	1.2	0.9	1.0	0.8	1.1	0.9

"COLLEGE PREP" SOCIAL STUDIES CLASSES

INCLUDES ADVANCED PLACEMENT CLASSES IN U.S. HISTORY, WORLD CIVILIZATION, AND EUROPEAN HISTORY; PLUS "COLLEGE PREP" CLASSES IN WORLD CIVILIZATION, U.S. HISTORY, AND WASHINGTON STATE HISTORY.



	ALL	AI	AN/AK	BL	HS	WH	REGS	LOWS	MALE	FEMALE
PREP SS ENROLLMENT	179	27	0	2	5	145	169	10	85	94
PROGRAM %		15.1	0.0	1.1	2.8	81.0	94.4	5.6	47.5	52.5
GROUP %	3.4	6.5	0.0	2.5	6.5	3.2	3.7	1.4	3.2	3.7
EQUITY INDEX:		1.9	0.0	0.7	1.9	0.9	1.1	0.4	0.9	1.1

SPECIAL PROGRAM ENROLLMENT

Rationale

Analyzing the enrollment in our special programs provides information on the extent to which we track or identify unique and separate programs for students in our districts. This indicator especially reflects on the extent to which we provide different tracks for students of different ethnic backgrounds and gender. Disproportional enrollment in a special program does not necessarily indict that program. If a student is appropriately placed in a remedial program, receives the remedial services necessary and is able to return successfully to the mainstream, the program has had a positive effect. If, however, students are not successful within the program and an ethnic group is disproportionally placed in that program, then our educational offerings are characterized by separate tracks and an appropriate education option has not been provided.

Definition/Data Sources/Formulas

An unduplicated count of students who have been enrolled in each of the special programs for any period of time during the past school year is used. While students enrolled in four different special programs are cited here, Chapter 1, Learning Assistance Program, Special Education (including Developmental Delay, SBD, SLD, MR Mild and Communication Disorders as reported for P223-H and Gifted Education Programs, this common definition has been used. The information is available from a variety of state and local district reports. Chapter 1, Learning Assistance Program and Special Education all require district report preparation under state guidelines and are available in each district. These various reports allow for the identification of students of each ethnic background and gender.

The attached graphic shows the percentage of each ethnic group served in each of our special programs (the "group percent"). This percentage is computed by dividing the number of students enrolled in that program for both genders and each ethnic group by the total number of students in the district who are in each ethnic group or gender. Provisions have been made to use only those grade level enrollments where the program is offered.

$$\frac{\text{\# subgroup student's in program}}{\text{\# subgroup students in district}} \times 100$$

Data Weaknesses

As indicated in the Rationale Section above, other information regarding the success of students who are enrolled in each program is necessary for a final judgement on the appropriateness or inappropriateness of tracking students into remedial or accelerated programs. An analysis of gain scores for students of each ethnic group who are enrolled in Chapter 1 would be one example of a program effectiveness measure.

Findings of the 1988/89 School Year

Before attempting a synthesis, at least three fundamental issues need to be addressed. First, there are basic difficulties with what some of the numbers represent. For example, should one divide December student counts by October student counts? If so, with what degree of confidence? One should ask do the numbers really mean what they seem to mean. Next, are disabilities, remedial needs, and brilliance normally distributed or might one group have greater or lesser needs than another group? How should one interpret comparative percentages in that regard? What constitutes over- or under- representation? In the SBD and SLD categories of special education, for example, are boys "over-identified" or should one naturally expect a skewed distribution? Finally, the minority populations vary from 7 percent to 31 percent across reporting districts. Since this synthesis considers ratios primarily, a caution must be added that widely varying denominators are at the base of those figures. Yet, even taking a conservative approach regarding these matters, one should not expect to observe a great deal of disparity. Nonetheless, that is what these data reveal in both gender and ethnicity.

The four programs of interest in the category of special program enrollment are:

A. Chapter 1
B. Learning Assistance Program (state-funded remediation program)
C. Special Education
D. Gifted Education

Six of the consortium's fourteen district members reported data. Those districts include Bellevue, Edmonds, Federal Way, Northshore, Shoreline and Tacoma. Of those, four reported Chapter I data, three reported learning assist program data and three reported gifted education data. In the area of special education all six districts reported data regarding program enrollment and Specific Learning Disability (SLD), Severe Behavior Disorders (SBD) and Mental Retardation (MR). However, only three districts reported about developmental delays and communication disorders. Additionally, the grade span and reporting period is not consistent among those reporting.

The six districts ranged in student population from 9,220 to 26,444 with an average population of 16,909. The ratio of minority to total population ranged from .07 to .31 with an average of .18. For the Chapter 1 program, the ratio of minorities participating in relation to the total minority population ranged from .09 to .22. There was a trend among the districts to have a smaller ratio of Asians and White and larger ratios of Blacks, Hispanics and Native Americans. However, Tacoma was the exception to that trend, having almost .22 across each ethnic category. Three districts reported gender figures which revealed slightly more males than females participating in the program.

For the Learning Assistant Program, three districts reported data. However, in one of these the program was so small, that if those data were used in comparison they would serve more to mask than to reveal. Therefore, LAP is considered from a two-district perspective. Within those two districts, again, there seems to be fewer Asians and Whites and more Blacks, Hispanics and Native Americans. In Tacoma, a ratio of .59 for Hispanics and .28 for Blacks is noted. Considering gender, there is about an even number of male and female participants.

Data from four districts were available in the gifted program area. Here the trend noted previously is reversed. In this program there are more Asians and Whites and fewer Blacks, Hispanics and Native Americans. When comparing combined minorities to Whites similar figures are observed, except in Tacoma where the balance favors Whites. There is a slight gender difference in favor of males in two of the reporting districts.

The program area of special education is particularly difficult to interpret because half of the districts reporting data did not include some of the categories under consideration. Further, inconsistency in grade span reported (e.g., K-6, K-8, K-12) and reporting periods considered (e.g., 1987-88, Dec. 1988, Jan. 1989 and Feb. 1989) also confound attempts at interpretation and inference. Nonetheless, the ratio of participating Whites to total Whites is higher than the ratio of combined minorities participating to total combined minorities. Even so, there is again a trend for more Blacks, Hispanics and Native Americans to participate than either Asians or Whites. The lowest ratio of participants in every district is the Asian ratio. The ratio of male to female in special education is overwhelming, ranging from a common two-to-one to as high as ten-to-one. Gender in special education is clearly an equity issue that deserves careful attention.

Having reviewed the data in this area, it is painfully apparent that much work needs to be done to improve the data collection process. Additionally, critical concepts and assumptions need further clarification and definition. Yet it is also apparent that these data support a call for a renewed effort. If, after improving the process, the data continue to show disparity and apparent inequities, a call for action would be in order. However, to act on the basis of these data alone is premature.

STANDARDIZED TESTING

Rationale

Recent revisions to most standardized tests have incorporated features that provide both norm referenced and objectives-based information. Major testing companies have designed their instruments to measure achievement in the basic skills areas commonly found in state and district curricula. These tests can provide information both about the ranking of students within a norm group and about instructional programs' strengths and weaknesses. The Metropolitan Achievement Test selected for use at grades four, eight and ten contains both features. With the administration of a student questionnaire, information becomes available on the performance of students in each ethnic group and for males and females. Information on several other questions relevant to excellence and equity is also available from the student questionnaire.

Definition / Data Sources / Formulas

The median national percentile rank IS reported not necessarily because it is the best metric to report on progress toward equity and excellence, but because this information that has been reported at the state level and for each member district during the past four years. The specific reports that will be prepared will include:

- o Median national percentile rank for grades four, eight and ten on total reading, total math, total language arts, and the total basic battery for each ethnic group.
- o Median national percentile rank for grades four, eight and ten for males and females on the above subtests.
- o An analysis of trends using median national percentile ranks for the past four years for each ethnic group in grades four, eight and ten on the MAT total basic battery score.

Data Weaknesses

Standardized tests are prepared by commercial testing companies and, while they are generally related to the curricula offered to students in K - 12 schools across the nation, specific alignment with a district instructional program is often lacking. Weaknesses have been pointed out in this test in the area of spelling and in mathematics, particularly in the area of problem-solving. A district that attempts to provide innovative programs runs the risk of developing courses of instruction not well aligned with standardized tests, which will result in students gaining considerable skills that are not reflected in increased standardized test performances.

Findings of the 1988-89 School Year

Standardized testing on the Metropolitan Achievement Test has been the most visible measure of school success for the past five years. The disaggregation of student performance by ethnic membership has been widely publicized in local district and in the media. In this paper we will not attempt to conduct a complete review, except to say that the pattern of performance that has been visible in Metropolitan Achievement Tests and in other standardized achievement tests is similar to the pattern that we have seen on other indicators within the profile. Black students, Hispanic students and Native American students score significantly below district averages. Asian students score at or above the district averages in many of the districts represented in the PSEC. This pattern appears to hold across each of the subtests that are reported.

CUMULATIVE GRADE POINT AVERAGE

RATIONALE

Course grades provide important measures of academic success. The grades students receive are more comprehensive indicators of success than achievement test scores. Being based upon professional judgement of the teacher, grades can reflect students' attitude and motivation, can be based on student performance, and can more easily reflect assessments of creativity and higher ordered thinking. Grades also can indicate student learning of the particular curriculum in each course in contrast to standardized test scores which indicate student learning of a sampling of curricula.

Furthermore, absolutely all students receive grades while the comprehensiveness of testing is never 100%.

Grade-point average, as an indicator, has the advantage of comparability between districts because of the state's standardized transcript, grade, and credit procedures.

DEFINITION / DATA SOURCES / FORMULA

Each semester, high school teachers submit grades for students in their classes. These grades and associated credits are added to high school transcripts and a cumulative grade point average (gpa) is computed. For high school gpa, all courses with grades of A, B, C, D, and E (or F), in courses for grades 9 and above are counted. Grade points and credits follow the state definitions and credit is earned for a minimum of 150 hours of instruction. Grade points are calculated on a four point scale with no plus or minus grades used. GPA is recalculated every semester.

For a subgroup, average GPA =

$$\frac{\text{sum of subgroup's high school student GPAs as of June 1988}}{\text{total number of students in the subgroup.}}$$

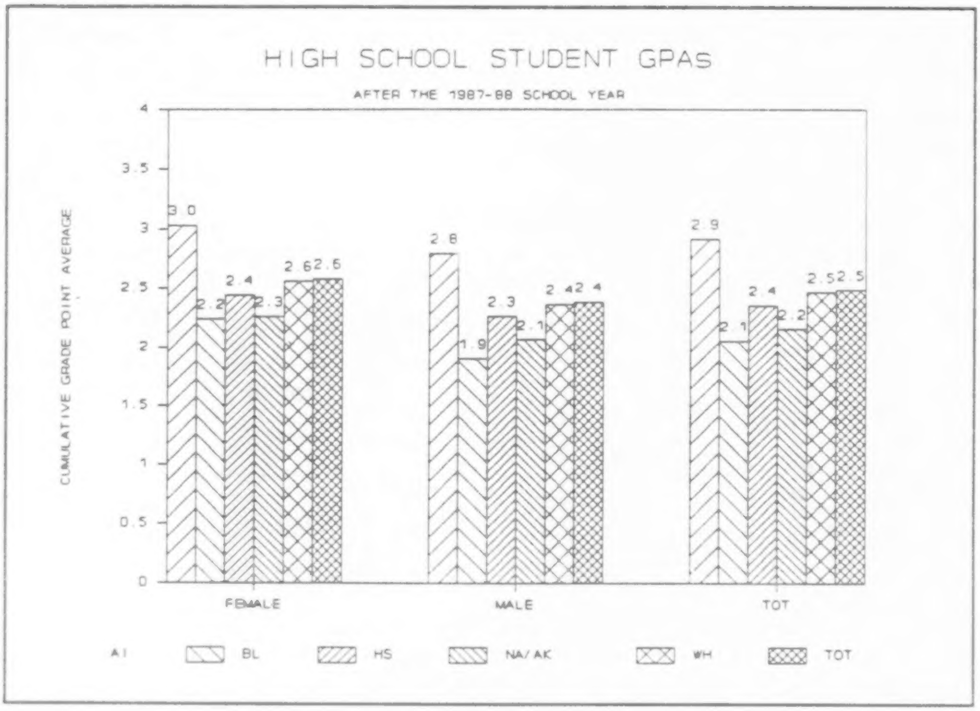
DATA WEAKNESSES

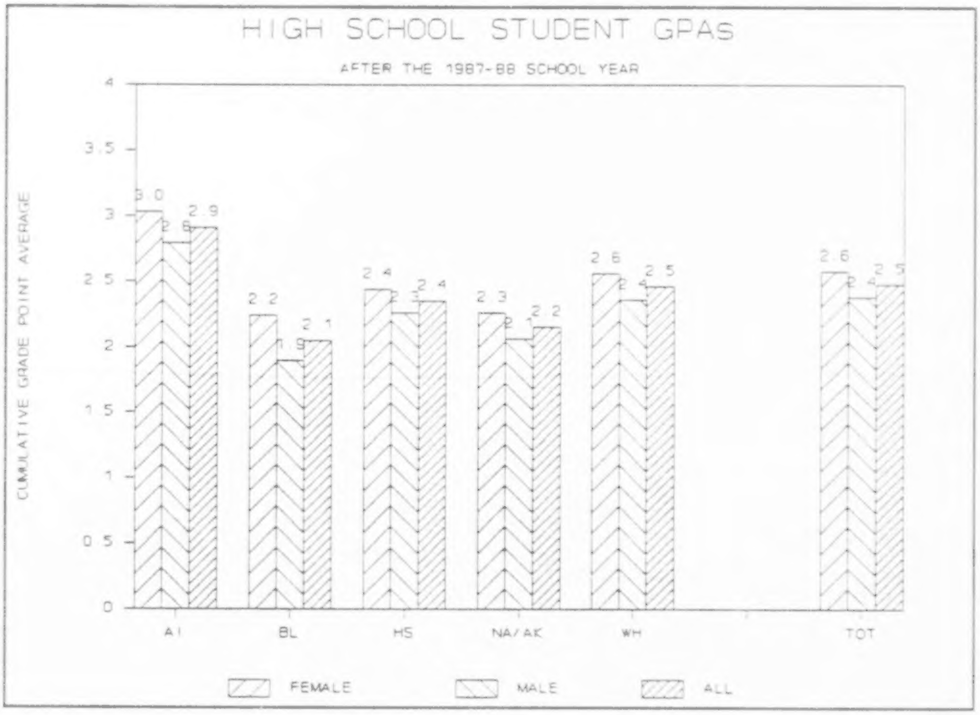
Schools and districts may vary in the degree to which they allow non-credit classes or pass-fail grading which would not be indicated in GPA. They may also vary in their grading standards or expectations for student performance. For these reasons, GPAs might not be equivalent between districts.

While teachers' grades are not limited to paper and pencil performances, their free form leaves them vulnerable to subjectivity, influences from personalities, and personality conflicts. In short, they are not necessarily indicative of student learning, although they have the potential of being the best indicators of that learning.

Student GPAs do reflect the completed high school classes on a transcript. When all GPAs are given equal weight for calculating subgroup averages (as they are here) then the freshmen's course grades are actually given more weight than seniors' grades.

Cumulative GPA is also not very sensitive to changes in grading practices, especially in the later high school years when students have accumulated two or more years of grades. More sensitive indicators might be semester grade point averages or the actual distribution of final grades awarded each term.


HIGH SCHOOL STUDENT GPAs
AFTER THE 1987-88 SCHOOL YEAR
CUMULATIVE GRADE POINT AVERAGE
0
0.5
1
1.5
2
2.5
3
3.5
4
3.0
2.2
2.4
2.3
2.6
2.6
2.8
1.9
2.3
2.1
2.4
2.4
2.9
2.1
2.4
2.2
2.5
2.5
FEMALE
MALE
TOT
AI
BL
HS
NA/AK
WH
TOT


HIGH SCHOOL STUDENT GPAs
AFTER THE 1987-88 SCHOOL YEAR
CUMULATIVE GRADE POINT AVERAGE
0
0.5
1
1.5
2
2.5
3
3.5
4
3.0
2.8
2.9
2.2
1.9
2.1
2.4
2.3
2.4
2.3
2.1
2.2
2.6
2.4
2.5
2.6
2.4
2.5
AI
BL
HS
NA/AK
WH
TOT
FEMALE
MALE
ALL

DATA TABLE: CUMULATIVE GPAs

	CUMULATIVE GRADE POINT AVERAGE					
	AI	BL	HS	NA/AK	WH	TOT
FEMALE	3.03	2.24	2.44	2.26	2.56	2.58
MALE	2.79	1.90	2.26	2.06	2.36	2.38
TOT	2.91	2.05	2.35	2.15	2.46	2.48
	NUMBER IN EACH SUBGROUP					
FEMALE	350	59	87	85	4733	5314
MALE	390	80	86	102	4748	5406

DESCRIPTIVE STATEMENTS.

- The average of Asian students'cumulative GPAs is significantly higher than other students' GPA averages.
- Black students and Native American students have lower average GPAs than other student groups.
- Female students' Average GPA is higher than the male students' average GPA overall and within each ethnic group. GPA differences are not as great between genders as they are between ethnic groups.

Findings of the 1988/89 School Year

Eight districts collected information for this indicator. The grades that students receive are predictable based on a student's ethnic background and gender. While some differences exist in the actual grades assigned and in the grade point averages of students across the eight districts, the rank order for grade point averages of each ethnic group and of males and females showed little variation.

- Asian/Pacific Islander students received the highest grades in each of the eight districts who supplied information. An average grade of 2.94 with a range from 2.69 to 3.14 across the districts was observed.
- The second highest grades were given to females who ranked second in the seven districts who supplied information for total females. The average grades given to females across the eight consortium districts was a 2.7.
- Black and Hispanic students received the lowest grades. The average GPA for Black students was a 2.12 and for Hispanic students 2.09. The district rank for each of these groups varied from fifth to seventh place for Black students and fifth to seventh place for Hispanic students, also.

Presented below is a table comparing grade points awarded to students of each of the districts with information available. At the request of one district, district names are replaced with numbers 1 through 8.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	PSEC AVG
AM IND/AL NAT									
GPA	2.05	2.15	1.98	2.41	2.1	1.94	2.37	1.75	2.0941
RANK	7	6	7	6	7	7	5	7	6.5
ASIAN									
GPA	3.14	2.91	2.98	3.03	2.99	2.75	3.05	2.69	2.9425
RANK	1	1	1	1	1	1	1	1	1
BLACK									
GPA	2.19	2.05	2.16	2.10	2.48	1.95	2.22	1.88	2.1276
RANK	5	7	6	7	6	6	7	6	6.25
HISPANIC									
GPA	2.18	2.35	2.46	2.44	2.51	2.22	2.33	2.04	2.3152
RANK	6	5	5	5	4	5	6	5	5.125
WHITE									
GPA	2.86	2.46	2.69	2.72	2.62	2.57	2.71	2.29	2.6149
RANK	3	3	3	3	3	2	3	3	2.875
MALE									
GPA	2.76	2.38	2.58	2.59	2.5	2.25	2.64	2.13	2.5106
RANK	4	4	4	4	5	4	4	4	4.125
FEMALE									
GPA	2.96	2.58	2.76	2.87	2.75	2.46	2.84	2.4	2.7368
RANK	2	2	2	2	2	3	2	2	2.125

WASHINGTON PRE-COLLEGE PROGRAM INFORMATION

Rationale

The Washington Pre-College Program is a cooperative, non-profit venture of the secondary schools, colleges and universities of the State of Washington. As a statewide program for the past thirty years, WPC provides services to guide high school students in career and educational planning and decision making. A wealth of trend data is therefore available about the characteristics of high school students who participate in WPC.

Definitions/Data Sources/Formulas

Washington Pre-College receives its information from students and schools.

- Students complete a registration form and pay a fee to participate in the WPC program.
- Schools provide transcripts for each student tested. WPC then summarizes the high school record through the junior year, providing GPA and credit summaries for English, mathematics, sciences, social studies, foreign languages, electives and overall.

Data selected for this report are trends in WPC verbal composite (VC) scores, WPC quantitative composite (QC) scores, and cumulative grade point averages through the junior year for annual testing from 1980 through 1987. Means, standard deviations and number of participants are listed in the tables.

Data Weaknesses

WPC participants are self-selected; the data include only students bound for in-state colleges. The data exclude the senior year academic record.

Very low numbers of ethnic minority students elect to take the WPC, so most district have to look at minority scores in the aggregate.

Findings of the 1988/89 School Year

Five districts reported Washington Pre-College information.

Overall GPAs of females have traditionally been higher than the GPAs of males. Females earn higher average course grades and achieve higher WPC test scores than males in English. Male students have traditionally obtained higher QC scores, but lower grades than their female counterparts. Mathematics GPAs for both males and females have been growing closer and are now virtually identical. Moreover, the gap between male and female QC scores has been narrowing in recent years.

Ethnic participants are less than 15 percent. Small increases have been observed over the past five years for Native American, Asian American and Hispanic students. For PSEC districts, Asian students appear to participate in the WPC program at a higher rate than other groups. In most districts, participation is less than ten students per year among Blacks, Hispanics and Native American/Alaska Natives. White students tend to dominate the mean scores and grades. Occasionally, Asian students significantly exceed mean scores and grades.

PROGRESS TOWARD GRADUATION - LOW CREDIT ACCUMULATION

Rationale

High school students spend four years earning credits toward graduation in order to complete their schooling successfully. Rate of credit earning thus provides one pragmatic way to identify students at risk. The high school students called "at risk" here are simply those who have not earned high school credits at a rate that will ensure they have enough to graduate with their class. These data are accurately maintained, often checked rigorously by students and building staff.

Definition/Data Sources/Formulas

The district's student information system contains data regarding the total number of credits each student has earned to date. Formulas for determining minimum number of credits-to-be-earned for satisfactory class standing will vary, depending upon the graduation requirements of a school district. For example:

CREDITS REQUIRED AT EACH GRADE LEVEL

GRADE LEVEL:	9	10	11	12
CREDIT NEED:	5	10	15	21

Students who proceed to the next grade level without the indicated minimum credits needed are at risk in their progress toward graduation.

Alternatively, in districts where students are assigned to grade level dependent on credits earned, credit accumulations must then be compared to a standard peer group, preferably the peers that entered ninth grade together. If necessary, credit accumulations can be reviewed for students who were 16, 17 and 18 years of age prior to September I in order to locate those with less than numbers of needed credit for sophomores, juniors and seniors respectively.

Data Weaknesses

The calculations here do not include information about which specific course requirements and competency requirements have or have not been met by students. A student could have sufficient credits without meeting all of the necessary graduation requirements. That student would not be identified as being at risk using this credit accumulation analysis.

Findings of the 1988-89 School Year

Two districts with dissimilar credit requirements for graduation reported that a total of 1484 students among 8051 enrolled tenth through twelfth graders were behind in appropriate credit accumulations for satisfactory class standing. In one case about 16 percent of the district's students appear not to be making minimum progress toward graduation. In the other case, about 20 percent of students are affected.

- Asian students appear to have fewer credit earning problems (lower percentage of students at risk) than any group.
- Black student groups tend to have higher rates of deficiency in credit earning than their district's average.
- Gender differences would appear to be significant in one district and not in the other; males may have more credit-earning problems than females.
- 31% of low income students (versus 19% of regular income students) are behind the credit-earning rate that is necessary for graduating on time from high school.

This indicator identifies a significant number of students who are at-risk and a significant degree of ethnic group and income group difference.

CRITERION REFERENCED TESTING

Rationale

The power of criterion referenced tests lies in the ability of the information gathered to reflect on the objectives we have chosen for our students to learn. Deciding what our students should learn and demonstrate on these tests can breathe life into the process of writing and teaching to SLOs. Measurement closely aligned with the teaching and learning process should also minimize the extent to which cultural or socio-economic backgrounds produce disproportionality in our measurements. In short, if these criterion referenced tests are well constructed, we should believe the results.

Data Sources, Definitions and Formulas

The percentage of items correct and the percentage of objectives mastered (learned) are usually computed by dividing total correct by total possible items on the test. Performance of students over time (three years) produces more convincing information and provides evidence of progress toward excellence. Gathering information for students of each ethnic group provides a message about equity. An analysis of the percentage of students scoring in each quartile of the test compared to the percentage that group represents in the total population tested will provide a more detailed measure of equity.

Data Weaknesses

Building reliable and valid tests is difficult work. One person's curriculum aligned criterion referenced test can be perceived as a complete mismatch for the teacher next door. Perhaps this reflects the variations in program implementation in our classrooms and schools. Patterns for the development of good tests with adequate input from teachers and measurement specialists are available and could become a focus for future PSEC efforts.

Comparisons are not available to national norms on locally developed tests, but trend analysis allows measures of progress toward both excellence and equity. Multiple years of information is necessary before information about program effectiveness becomes available.

Criterion-referenced test results will never be comparable across districts and so should be less useful than other indicators for looking at comparative excellence. This characteristic, however, should do nothing to diminish the indicator's value for looking at equity.

The difficulty of disaggregating district level tests is significant. Developing a process to identify student ethnicity and gender as part of the answer document or to develop test scoring systems that can communicate with student record files so that gender and ethnicity information may be added to test information is one necessary step before this data can be disaggregated in those districts who give criterion referenced tests.

Findings of the 1988/89 School Year

Only one consortium district was able to develop disaggregated information for criterion referenced tests. In the single district where this information was available, the pattern identified in standardized testing continued.

- Black and Hispanic subtest scores continue to be the lowest of the ethnic comparisons. Although the number of students involved in these comparisons was small, the consistency across test administered at grades 3, 5 and 9 in mathematics added to the confidence that these are accurate representations of student scores.
- In the disaggregation for males and females within each ethnic group, Black males performed higher than Black females on this particular indicator, a finding that reversed information on several other indicators.
- Competency testing in the area of reading followed a fairly similar pattern with Black and Hispanic scores consistently the lowest in the district.

COMMUNICATION SKILLS ASSESSMENT

Rationale

Assessment of students written communication skills provides valuable information about the effectiveness of writing instruction in language arts. In Bellevue, each student is required to demonstrate his or her competence on ten measures of good writing as one requirement for high school graduation. Writing assessment is conducted for all students enrolled in grades five, eight and ten. Gathering this information to reflect student performance at each school and each grade level provides a measure of school effectiveness.

Definitions/Data Sources/Formulas

The tables and charts on the following pages presents student performance for each of the past three years on the Bellevue Competence Test in writing. Both the average score for students and the total percentage of students who demonstrated competence on this assessment are available. The percent of students who achieved competence in each ethnic group and for each gender are presented for the tenth grade test administration. These reports contain information about program excellence when reviewing information about changes in the percentages of students achieving competency or changes in average scores. Different score values for students of different ethnic backgrounds provides information about equity in the writing instructional process.

Data Weakness

The assessment of student writing produces less reliable scores than those coming from either standardized or criterion referenced tests. Comparisons of average scores should be done with caution. The percent of students achieving competence is the most reliable score value for program evaluation.

Findings of the 1988/89 School Year

Communication Skills Assessment information is most often defined by writing samples prepared by students. Only one district was able to report this information pointing out the difficulty of conducting a study involving disaggregation of information for ethnicity and for gender. Information from the single district that conducted a disaggregation was completed at the high school tenth grade level. This information indicated that Hispanic students were the lowest performing population, followed by Asian and Black students.

The measurements taken were the percentage of students who achieved competency in writing on the initial writing sample taken in May of students' tenth grade year. Differences between ethnic groups were as high as 28 percent with just 23 percent of Hispanic students achieving competency, while 51 percent of white students were rated competent. A significant gender difference also was evident with 41 percent of males tested judged competent, compared to 59 percent of the females. Preliminary inspection indicates that gender differences may hold across ethnic groups at least at some grade levels.

STUDENT OPINION ABOUT SELF

Rationale

Student self report data have been proven useful for identifying potential dropouts from high school. According to research literature on the topic of high school dropouts and student success in high school, students' educational goals, views of themselves as learners, and self esteem play fundamental roles. Assessment of these is key to understanding student groups and individual and group differences.

The survey designed for gathering these data can also ask for self-report information about discipline experiences, attendance, school and community activities and high school plans. These all seem important to the study of equity and excellence in the schools, yet most districts have incurred difficulty trying to gather the information in a uniform, reliable manner. If verified as a reliable information resource (and analyses to date cause this district to have confidence in the data), this type of survey **may prove to be the most reliable and uncomplicated means for gathering student information for profiling.**

Definition/Data Sources/Formula

High school students' opinions have been gathered for this analysis via a locally edited version of the U.S. Department of Education's High School and Beyond Survey.

The use of student names and identification numbers on the survey answer sheets allows survey answers to be merged with information from the district student data base which provides information such as ethnic category, free lunch eligibility and actual grade point average.

For each item or construct, the indicator is the percentage of each responding subgroup giving a particular answer. Tables and graphs represent, for example, the percentage of each subgroup with plans to go on to college, the percentage who have experienced disciplinary actions during the last school year, and so on.

Data Weaknesses

Completion of the survey is supervised in the classroom by the classroom teachers. This may detract from the validity due to lack of uniform procedures.

Some people have a negative attitude toward the collection of data that they perceive as personal. This may produce a proclivity to respond in less than candid ways.

Since the data are self-reported, with no incentive for honesty and candor, there is no assurance that the responses are true. (However, validity tests performed on parts of the data suggest about 98 percent accuracy).

Certain items in the survey may be sensitive to the current mood of the student as opposed to the student's general attitude of self esteem.

Findings of the 1988/89 School Year

Statistical analyses of High School and Beyond Survey responses from 5,000 students in one school district showed many predictable and other surprising significant differences ($p<.01$) between subgroups. For example, gender group differences are apparent:

- A significantly higher proportion of females say it is true that "At times I am no good at all." Ethnic group differences on this item are not significant.
- Males are significantly more satisfied with themselves than females.
- Males feel significantly more confident that they can make their own plans work than females.
- More males than females report NO involvement in high school activities.
- Females are absent without being ill more often than males.
- Males and females did not differ on the composite measure of self esteem.
- Nearly twice the percentage of males as females report discipline problems.

Results also reveal considerable ethnic group differences:

- More Native American students report not being involved in high school or community activities.
- Native American students showed the lowest self esteem and Hispanic students showed the highest level of self esteem.
- Hispanics and Native Americans report a significantly higher rate of being absent from school when they were not ill.
- Thirty five percent of Native Americans versus 17 percent of Asians report having discipline problems in school. The overall percentage of students reporting having discipline problems in school during the last school year is 18 percent, the exact same percentage as was determined from central office records.
- Although Asians report higher grades, higher aspirations and lower percentages of disciplinary actions, they answer some self esteem items less positively than other groups of students. ("At times I think I am no good at all" and "I am satisfied with myself").

DISCIPLINE

Rationale

Formal disciplinary activities are usually defined by suspensions from school or suspensions from selected classes in school. These formal disciplinary actions follow either severe misbehavior or a series of violations of school rules. The extent to which suspension occurs in a school is often considered to be one indicator of school climate. Positive classroom management can usually lead to a lessened need for formal disciplinary activities. To the extent this is true, the formal discipline information contained on the following pages is a valid indicator of program quality.

Definitions/Data Sources/Formulas

The information presented in most district reports includes the record of formal discipline activities for each of our 6 - 12 schools. The table presents both the number of students receiving formal discipline and the number of disciplinary incidents. The information is gathered by administrators in each school and is reported yearly to the district affirmative action officer. The table and graph contain a simple count and a percentage comparison of disciplinary activities given to males and females of each ethnic group compared to the percentage that group represents of our total 6 - 12 population. To the extent that discipline percentages differ from enrollment percentages, disproportionality exists. To the extent total disciplinary incidents change over time, school climate has changed.

Data Weaknesses

As staff and administration changes at our schools, the tactics of behavior management also change. To the extent these changes impact formal discipline, the reliability of these counts is lessened. New policies or procedures could negatively affect this measurement as an indicator of school quality.

Districts collect information about disciplinary actions taken, not about infractions of rules. Actions are influenced by policy and personality and are not meted out consistently between schools or districts for the same infractions.

The concept of equity is certainly important here if different subgroups of students receive different repercussions for the same infractions. The consortium problem is that we now have no way of discerning the real differences -- if there are differences.

Findings of the 1988-89 School Year

Because of a complete lack of quality data, no district level synthesis information was available for this report.

DEVELOPMENT OF A 50-STATE SYSTEM OF EDUCATIONAL INDICATORS: ISSUES OF DESIGN, IMPLEMENTATION, AND USE*

Rolf K. Blank
Council of Chief State School Officers

Local school districts, state departments of education, and federal agencies, such as the Department of Education and the National Science Foundation, have been active in trying to improve the quality and usefulness of data and information on education. Many of these efforts have been aimed at developing and implementing a system of educational indicators that would regularly inform policy-makers, the public, educators, and researchers about the extent of progress in improving education in our schools (Oakes, 1986; National Science Board, 1987; National Center for Education Statistics, 1987; Phi Delta Kappan, 1988; National Governors' Association, 1986; Council of Chief State School Officers, 1988a).

States have become very active in developing indicators of education performance to assess effects of the wide range of state education reforms. In the 1980's, states have increased graduation requirements, revised elementary and secondary curricula, upgraded teacher certification requirements, created incentives to attract and retain teachers, and developed and revised state assessment tests (Blank & Espenshade, 1988). The state education commissioners and their staffs have been working to track the effects of these reforms and to ensure that methods are available for reporting on the progress of states in improving education (U.S. Department of Education/OERI, 1988).

In 1984, the Council of Chief State School Officers (CCSSO) made a fundamental shift in its policy on the role of states in educational assessment and evaluation. The state commissioners adopted a far-reaching position on the responsibility of states to collect and report education data that have high technical quality and that will provide useful state-to-state comparisons. CCSSO established the State Education Assessment Center to coordinate the development, analysis, and use of state-level data and charged the Center with implementing an education indicators model to report state-by-state data.

---

*This paper includes data collected by state departments of education that was obtained through the cooperation of the State Network on Science/Mathematics Indicators, comprised of professional staff in each state department of education. Research support was provided through a grant from the National Science Foundation, Office of Studies and Program Assessment, Science and Engineering Education.

This paper reports on the process of developing a state-by-state indicators system used by CCSSO and outlines how some of the indicators can be used. There are three sections. The first section describes the CCSSO model for state education indicators. Then, two projects within the State Education Assessment Center are summarized to illustrate the process of developing comparative state-by-state indicators which use different sources of data: a) the State Science/Math Indicators Project, an effort to identify comparative indicators based on data from state information systems; and b) the National Assessment Planning Project, an effort involving states and national education organizations to prepare for state-by-state comparisons of student achievement.

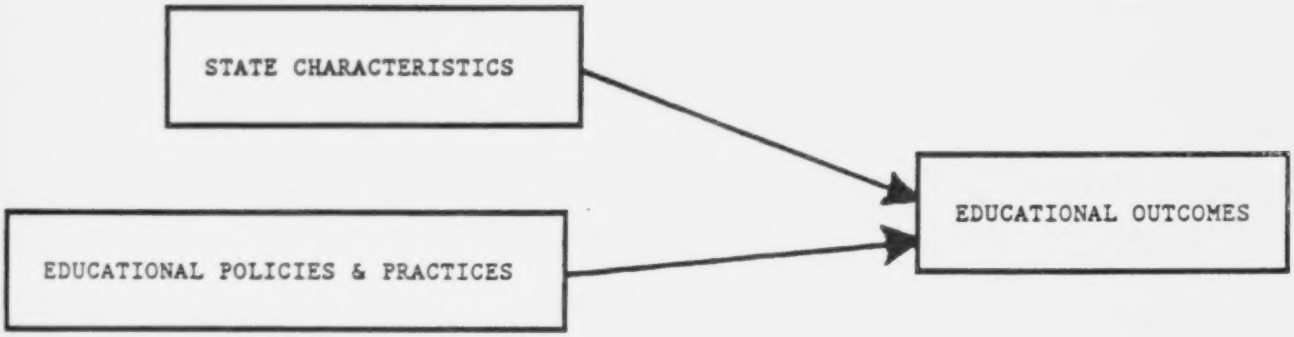
MODEL FOR STATE EDUCATION INDICATORS

The CCSSO indicators model has three components: a) educational outcomes, b) state educational policies and practices, and c) state background characteristics. The goal of the model is to relate outcomes to policies and practices, accounting for state demographic and fiscal characteristics that determine, to some extent, what schools can accomplish. This gives the indicators explanatory power that they would not have as individual variables, because the scheme is intended to model, based on research, educational inputs and outputs and the relationship between them.

The CCSSO model, shown in Figure 1, was initially recommended by a committee of state superintendents and their staff (CCSSO, 1985). The committee also established three criteria for selecting indicators: a) importance/utility for states, b) technical quality of data, and c) feasibility of implementing on a state-by-state basis.

Figure 1

CCSSO EDUCATION INDICATORS MODEL & SCIENCE/MATH INDICATORS



EDUCATIONAL POLICIES & PRACTICES

Current Report: State Policies

- o Graduation requirements
- o Teacher preparation/certification
- o School participation

Developing: State-by-State Data

- o Allocated time in elementary subjects
- o Science & Math Indicators:
  - - Secondary course enrollments
  - - Teacher assignments
  - - Assignments by certification
- o Teacher work force
- o Expenditures

STATE CHARACTERISTICS (Current)

- o School System Demographics
- o Population Characteristics
- o Resources
- o Student needs

EDUCATIONAL OUTCOMES (Developing)

- o Graduation rate & dropout rate
- o Student achievement
- o Post high school status

Since 1985 the State Education Assessment Center has been implementing a system for reporting state-by-state data on the indicators. As shown in Figure 1, some of the indicators are now in place and being reported on an annual basis (CCSSO, 1988), while others, such as student achievement and dropout rates, are being prepared. The Assessment Center has three principal roles in implementing the indicators model:

1. Developing a consensus among states on the specific measures that will be used for each indicator;
2. Providing technical assistance to states to improve their systems of assessment, accountability, and data collection; and
3. Working with federal agencies and others collecting educational data to ensure that valid state-by-state comparisons can be collected and reported.

The Assessment Center works closely with state education specialists through planning groups or task forces. A state task force could be comprised of specialists responsible for state information systems, assessment programs, finance data, or curriculum, with the composition varying according to the specific type of indicator being planned. For each category of indicators, state representatives provide recommendations on indicator development and report on what can be done by each state department of education. The states then collaborate with Assessment Center staff in a series of steps to collect and report comparative state data, including:

- Establishing common definitions and standards for data collection;
- Surveying states to identify current state indicators and data;
- Analyzing differences between states in definitions and procedures;
- Recommending adjustments to standardize data across states; and
- Providing technical assistance to states for improving data systems and reporting data.

SCIENCE/MATHEMATICS INDICATORS PROJECT: Process of Developing Comparative Indicators with State-Collected Data

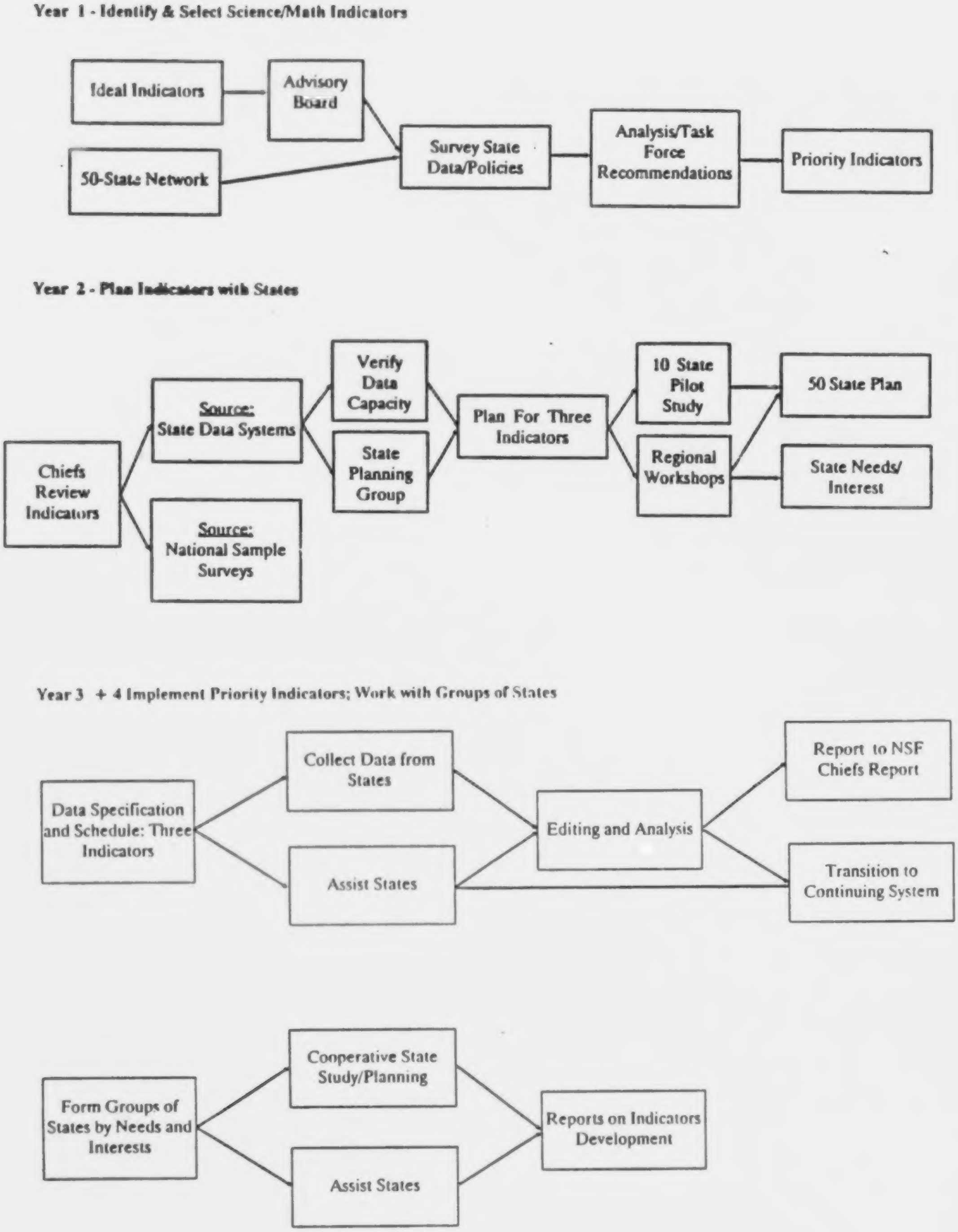
To develop the indicators in the CCSSO model, one approach was to gain support from federal agencies that have a similar interest in educational indicators. CCSSO found that the National Science Foundation's (NSF) interest in improving the nation's capacity for monitoring science and math education coincided with the states' interest in comparative indicators at the subject level in areas such as student achievement, student participation, teacher qualifications, and resources.

With support of the National Science Foundation, the Assessment Center began a project in 1986 to develop indicators of the condition of science and mathematics education at elementary and secondary levels. The goals of the project, now in its third year, are to: a) improve the quality and usefulness of data on science and mathematics education to assist state policy-makers and program managers in making more informed decisions, and b) to develop a system of indicators that provides the capacity for state-to-state comparisons of science and mathematics education as well as national indicators to assess the condition of education in these subjects. The project has been conducted through planned steps, as displayed in Figure 2 and summarized as follows.

1. Conceptual Framework for State Science/Math Indicators

Initially, a conceptual framework for state-by-state science and math indicators was designed. A paper was drafted (Blank, 1986) based on the CCSSO model as well as recent recommendations on indicators of science and mathematics education, such as NSF-supported studies by the National Academy of Sciences and the Rand Corporation (Murnane & Raizen, 1988; Shavelson, et al., 1987). Then, the draft was reviewed and revised by an advisory board comprised of scientists, mathematicians, state education specialists, and education researchers. The framework identified six areas of desired, or "ideal," indicators for science and

Figure 2


Year 1 - Identify & Select Science/Math Indicators
Ideal Indicators
Advisory Board
50-State Network
Survey State Data/Policies
Analysis/Task Force Recommendations
Priority Indicators
Year 2 - Plan Indicators with States
Chiefs Review Indicators
Source: State Data Systems
Source: National Sample Surveys
Verify Data Capacity
State Planning Group
Plan For Three Indicators
10 State Pilot Study
Regional Workshops
50 State Plan
State Needs/ Interest
Year 3 + 4 Implement Priority Indicators; Work with Groups of States
Data Specification and Schedule: Three Indicators
Collect Data from States
Assist States
Editing and Analysis
Report to NSF Chiefs Report
Transition to Continuing System
Form Groups of States by Needs and Interests
Cooperative State Study/Planning
Assist States
Reports on Indicators Development

math: student outcomes, instructional time/enrollment, curriculum content, teacher quality, school conditions, and resources. The framework also outlined the rationale for improving indicators of science and mathematics based on needs of education policy-makers and program managers at national and state levels. The key elements of the rationale are outlined in Figure 3.

2. Survey with States to Identify Commonalities and Differences in Indicators

Based on the conceptual framework, the project surveyed the state departments of education to determine which of the ideal indicators are available through state assessment programs, information systems, or special research studies. Each state commissioner was asked to select three state education staff, including specialists in science and mathematics curriculum, assessment, and information systems, and these staff were asked to complete the survey and to serve on a continuing project network. One reason for such a network is to have staff from different divisions of state agencies discuss and share their ideas about the needs for and uses of indicators of science and mathematics education, as well as to ensure that state data provide valid, reliable measures of the indicators.

The survey results were used for several kinds of analyses. First, the availability of desired indicators among the 50 states was determined (Blank & Espenshade, 1988). Second, for states that have data on a given indicator, the specific collection forms and data systems were analyzed to determine whether data could be reported in a common way with only small changes, adaptations, or additions. Third, if a given indicator was selected, the number of states that would need to add new data elements or forms was determined.

3. Task Force Recommendations for Priority State-by-State Indicators

The project convened a task force comprised of state participants in the

Figure 3

WHAT IS THE PROBLEM?

WHY EDUCATIONAL INDICATORS?

I. NATIONAL

- ECONOMIC COMPETITION (INTERNATIONAL)
- SCIENCE/MATH LITERACY
- CAPACITY FOR MEASURING PROGRESS
    - OUTCOMES
    - PARTICIPATION
- RELATE TO POLICIES

II. STATES

- STATE POLICIES/REFORMS
- USEFUL STATE-TO-STATE COMPARISONS
- IDENTIFY/ANALYZE PROBLEMS
- PLANNING & ASSISTANCE

project network and experts on science/math indicators to analyze the survey results and recommend a set of priority state-by-state indicators for science/math education. The three CCSSO criteria for selecting indicators were applied, i.e., importance/utility, quality, and feasibility. In order for an indicator to be selected it needed to receive a high rating by the task force as being important and useful for tracking state differences and trends in science and math, available state data needed to have high technical quality, and data collection and reporting on a state-by-state basis needed to be feasible.

The recommendations of the task force for "priority indicators" are summarized in Figure 4. Two types of data sources were identified for producing the priority indicators: national surveys with representative samples by state (NAEP and Schools and Staffing Survey) and universe data collection from state administrative records, such as teacher personnel or school reporting forms.

4. Plan for Reporting Comparative State Data

To use state-collected data, the CCSSO project worked with a second task force of state specialists in science, mathematics, and information systems to develop a reporting system and reporting format that would produce comparable data for the selected indicators. For example, for three secondary-level indicators, the reporting plan includes a taxonomy of course and teacher assignment categories, a "cross-walk" that relates each state's course and teacher data codes to the common reporting format, and definitions of the science and math categories in the taxonomy (CCSSO/SEAC, 1988).

5. Regional Workshops on Implementing Data Plan and Uses of Indicators

To assist states in implementing the science/math indicators, a series of regional workshops were conducted across the country with the state education staff in the project network. About 110 state education staff participated in one of the five regional meetings. The workshops provided a way for states to

Figure 4

Council of Chief State School Officers

PRIORITY STATE-BY-STATE INDICATORS OF SCIENCE & MATH*

INDICATOR	DATA SOURCE
Student Outcomes	
STUDENT ACHIEVEMENT	NAEP (1990)
STUDENT ATTITUDES/INTENTIONS	NAEP (1990)
Instructional Time/Enrollment	
GRADES 7-12 COURSE ENROLLMENT	State Data (CCSSO**)
ELEMENTARY MINUTES PER WEEK	Schools/Staffing Survey (CES, Spring 1988)
Curriculum Content	
STUDENTS' "OPPORTUNITY-TO-LEARN"	NAEP (1990)
School Conditions	
CLASS SIZE by Subject/Course	Schools/Staffing Survey and State Data (Available in some states, not currently requested by CCSSO)
NO. of COURSE PREPARATIONS PER TEACHER	
COURSE OFFERINGS PER SCHOOL	
Teachers	
COURSES/CREDITS IN SCIENCE/MATH	Schools/Staffing Survey (CES, Spring 1988)
TEACHING ASSIGNMENTS By Subject/Field By Age, Gender, Race/Ethnicity	State Data (CCSSO**)
TEACHING ASSIGNMENTS BY CERTIFICATION FIELD/SUBJECT (Number of Uncertified Teachers)	State Data (CCSSO**)
Equity	
GENDER AND RACE/ETHNICITY by Indicator	State Data (CCSSO**) (where available)

---

* Science/Math Indicators Project, CCSSO State Education Assessment Center, February, 1988. Priority indicators were recommended by Project Task Force (September, 1987) and reviewed by CEIS (September, 1987) and CCSSO (November, 1987).

** To be reported to CCSSO on a state-by-state basis through the Science/Math Indicators Project.

individually and collectively respond to the reporting plan and specifications. They had the opportunity to discuss the assumptions and rationale used by the task force and staff. Second, state participants discussed the uses of science/math indicators for policy and program decisions in states and how state data systems can be improved. Third, recommendations were made on additional indicators in science and mathematics that states would like to see developed.

For example, the discussion of uses of indicators identified several questions of interest for state policy analysis and program planning that could be answered with the secondary-level science/math indicators to be reported by states:

Indicator: Enrollment in Science/Math Courses in Grades 7-12

- What types of courses are students taking to meet state science and math graduation requirements?
- How does course-taking in science/math differ in LEAs and schools with different characteristics?
- What is the trend in student enrollments in science and math courses from one year to the next? What is the rate of attrition from science and math (by examining enrollments over time)?

Indicator: Teacher Assignments in Science/Math

- What is the age of teachers in science/math, particularly the numbers of new and older teachers? What is the race/ethnicity and sex distribution of science/math teachers?
- How many different course preparations do most science/math teachers have?

Indicator: Assignments by Certification

- What are the state-by-state differences in shortages of science and math teachers, as indicated by out-of-field teaching?
- How do the data on current teachers show need for funding support of incentive programs, such as scholarships or retraining?

At the regional workshops, states also discussed how science/math indicators can be used within states to examine educational progress according to characteristics of districts and schools, such as size, student composition, and local policies.

6. Pilot Study of State Data Reporting on Science/Math

In the spring of 1988, ten states participated in a pilot study to test the common reporting format for the three indicators to be based on state-collected data. These data provided an initial analysis of the applicability of the plan for data reporting and identified any problems states had in using the formats, categories, or definitions. The analysis of the pilot study was used to revise the instructions and forms for the initial year of state reporting during the 1988-89 school year (CCSSO/SEAC, 1988).

Examples of the data from the pilot study are reported in Tables 1-4. Tables 1 and 2 show individual state data according to the secondary course enrollment taxonomies for science (Table 1) and mathematics (Table 2). The data from these two states also provide breakdowns of course enrollments by gender.

Several basic principles were used in developing the taxonomy for reporting course enrollments. First, the course categories are intended as measures of the relative level of student progress in science and mathematics, for example, the percent of students taking first-year general biology vs. the percent taking first-year applied biology. They are not intended to provide comparisons across states in actual course content. However, the taxonomy does include category definitions that states can use in checking their data categories against the common reporting format. Second, the taxonomies for both science and math are designed to be forward-looking in anticipating the direction of curriculum development, such as integrated math for different high school grade levels, and the increase in science and math courses, e.g., "applied" courses, with the increase in graduation requirements.

Tables 3 and 4 show examples of the pilot study data on teacher characteristics. The teacher data were reported by teaching subject (e.g., math, biology, chemistry, etc.) and by teachers with primary assignment (50% or more of their time) in a specific subject versus those with secondary assignment (less than 50% of their time) in the subject. In Table 3, the teacher assignments for those 50% or more (primary assignment) in biology are cross-tabulated with

Table 1

SECONDARY COURSE ENROLLMENT - SCIENCE

STATE 1
(1987-88)

Science Courses	Total Enrollment	Boys	Girls
Grades 7-8			
General Science	4,973	N/A	N/A
Life Science	44,848	N/A	N/A
Earth Science	44,851	N/A	N/A
Physical Science	579	N/A	N/A
Other Science, 7-8	541	N/A	N/A
Grades 9-12			
Biology, 1st Year			
Biology, 1st Year, General	47,282	23,736	23,546
Biology, 1st Year, Applied			
Biology, 2nd Year, Advanced Placement	1,145	609	536
Biology, 2nd Year, Other Advanced	3,263	1,351	1,912
Chemistry, 1st Year			
Chemistry, 1st Year, General	18,774	8,805	9,969
Chemistry, 1st Year, Applied			
Chemistry, 2nd Year, Advanced Placement	Not Offered *		
Chemistry, 2nd Year, Other Advanced	1,009	564	445
Physics, 1st Year	4,800	2,861	1,939
Physics, 1st Year, General	4,508	2,642	1,866
Physics, 1st Year, Applied	292	219	73
Physics, 2nd Year, Advanced Placement	Not Offered *		
Physics, 2nd Year, Other Advanced			
Earth Science, 1st Year			
Earth Science, 1st Year, General			
Earth Science, 1st Year, Applied			
General Science	15,878	8,939	6,939
Physical Science	33,250	16,791	16,459
Other Science, 9-12	2,526	1,465	1,061

---

* These courses were not offered during the 1987-88 school year.

Table 2

SECONDARY COURSE ENROLLMENT - MATHEMATICS

STATE 2
(1987-88)

Mathematics Courses	Total Enrollment	Boys	Girls
Grades 7-8			
Math, Grade 7	50,608	25,778	24,830
Math, Grade 7, Accelerated (Pre-Algebra)	—	—	—
Math, Grade 8	50,577	25,906	24,671
Math Grade 8, Accelerated (Algebra 1)	5,035	2,420	2,615
Grades 9-12			
Review Math			
Level 1 (General, Remedial)	24,212	13,339	10,873
Level 2 (Consumer, Vocational, Applied)	18,735	10,011	8,724
Level 3 (General Math 3)	—	—	—
Level 4 (General Math 4)	—	—	—
Informal Math			
Level 1 (Pre-Algebra, Basic, Algebra 1(A)*)	18,530	9,473	9,057
Level 2 (Basic Geometry, Practical)	—	—	—
Level 3 (Basic Algebra 2)	—	—	—
Formal Math			
Level 1 (Algebra 1, Algebra 1(B)**)	59,164	29,741	29,423
Level 2 (Geometry, Plane, Solid)	45,092	22,000	23,092
Level 3 (Algebra 2, Intermediate)	28,541	14,073	14,468
Level 4 (Trigonometry, Advanced Math)	28,744	15,322	13,422
Level 5 (Calculus)	2,550	1,478	1,072
Level 5, Advanced Placement (Calculus)	—	—	—
Other Mathematics, 9-12	—	—	—
Computer Science Courses			
Grades 7-8			
Computer Science/Programming	11,725	6,079	5,646
Grades 9-12			
Computer Science/Programming I	31,906	18,135	13,771
Advanced Computer Science/Programming II	12,135	6,530	5,605

---

* First year of a two-year course sequence

** Second year of a two-year course sequence

Table 3

TEACHERS ASSIGNED 50% OR MORE BIOLOGY (GRADES 9-12)

BY AGE, SEX, RACE/ETHNICITY

		Age		Sex		Race/Ethnicity				
STATE	No. Teachers 50% or More Biology	Under Age 30	Over Age 50	M	F	H	W	B	A	I
State 1	589	43 (7%)	72 (12%)	226	363		468	117	2	2
State 2	431	76 (18%)	44 (10%)	168	263		332	99		
State 3	1,485	82 (6%)	6 (0.4%)	1,074	411	3	1,440	38	2	2
State 4	282	27 (10%)	47 (17%)	156	126	1	270	9	1	1
State 5	829	43 (5%)	228 (28%)	698	131	3	816	8	1	1
State 6	68	5 (7%)	11 (16%)	58	10		68			
State 7	283	33 (12%)	35 (12%)	128	155		249	33		

Table 4

TEACHERS IN MATHEMATICS (GRADES 9-12) BY CERTIFICATION STATUS *

STATE	Total	Assigned 50% or More			Assigned Less than 50%		
		Total	Math Certified	Out of Field	Total	Math Certified	Out of Field
State 1	1,580	1,216	1,190	26 (2%)	364	327	37 (10%)
State 2	1,933	1,625	1,500	125 (1%)	308	101	207 (67%)
State 3	5,522	5,353	5,281	72 (1%)	169	135	34 (20%)
State 4	1,719	1,404	1,353	51 (4%)	315	132	183 (58%)
State 5	472	292	292	0	180	180	0

---

* For 1988-89, data will be displayed with each state's certification requirements.

teacher age, sex, and race/ethnicity. The two age categories indicate the numbers of younger teachers and older teachers for each state. The age proportions probably follow the age profile of all teachers in a state, but they also point out potential teacher shortages in science/math. For example, State 5 has a quarter to a third of teachers with primary assignments in science/math who are over age 50. The data on gender indicate that States 1, 2, and 7 have more female than male science and math teachers, while the other states have more male teachers. All of the states appear to have under-representation of minority teachers given their student populations. It is likely that periodic collection of these data would produce very useful trend analyses of science/math teacher characteristics by state.

Table 4 illustrates state data on the indicator of teacher assignments by certification status, in this case for mathematics. The indicator is based on a cross-tabulation of teacher assignments (above or below 50% time) in mathematics, biology, chemistry, physics, earth science, and physical science by their certification status in the specific subject. These data provide a method of identifying the number of teachers in each field that are teaching "out of field" (which includes certified teachers who are assigned to a field in which they are not certified, and emergency, temporary, or provisionally certified teachers). For example, 4 percent of teachers in State 4 assigned 50% or more in math are not certified in math, while 58% of the teachers assigned less than 50% in math are not certified in math.

7. Implementation of Indicators System.

In the third year of the project (1988-89 school year), states began reporting data to CCSSO on three secondary-level indicators. The reporting system was phased-in with those states that had the data available through state information systems (e.g., 32 states had data on all three secondary indicators). The project assisted other states in developing their state data capacity. Data on other priority science/math indicators were obtained from state-by-state analysis of results of the NCES Schools and Staffing Survey.

In the 1989-90 school year, the project will coordinate the full implementation of state data reporting on the three secondary science/math indicators. At the same time, CCSSO will be working with the National Science Foundation to develop and implement a plan for continuing state-by-state reporting of the science/math indicators on a biannual basis. This schedule will allow the state data to be reported with the results of the state-by-state NAEP. The data collection programs of the National Center for Education Statistics are currently being considered as a way of collecting and reporting the data from states.

During the current year, the project is working with small groups of states on developing additional indicators. The project organized a meeting of 10 state science supervisors to identify and plan indicators of elementary science in November 1988. During the next year the project will be working with states to determine how data could be collected and reported to provide valid, state-by-state comparisons on elementary science. From the recommendation of the science supervisors, the project planned and organized a conference on "Alternative Methods of Science Assessment" in January 1989 which was attended by 60 state assessment directors and state science supervisors from 39 states (Blank & Selden, 1989). Presentations on recent developments in science assessment provided the basis for states to consider ways of improving state assessment programs.

NATIONAL ASSESSMENT PLANNING PROJECT: Developing a Basis for State-by-State Comparisons of Student Achievement in Mathematics

In 1987-88, the CCSSO State Education Assessment Center led a consortium of 18 national education organizations and state education representatives in developing a plan for expanding the National Assessment of Educational Progress (NAEP) to include state-by-state comparisons. For twenty years NAEP served as "The Nation's Report Card" in providing the only comprehensive, longitudinal data on what students in the U.S. know and can do in various subject areas. NAEP was based on a national representative sample of 25,000 students at each of three grade levels. For the 1990 assessment, states will be able to participate in a state-by-state assessment in mathematics, and in 1992 in reading.

The National Assessment Planning Project led by CCSSO filled a specific need of the states and the U.S. Department of Education, which was to obtain a consensus among the states on how to proceed with designing a national assessment in mathematics that would provide valid, useful state-to-state comparisons. There were two major questions of concern to state and local educators and policy-makers. First, there was a question about whether the educational objectives (intended knowledge, skills) upon which the assessment would be based reflects what states have set as their educational goals and objectives. Second, there was concern about how the comparative data would be analyzed and reported so that states would be compared in a fair and meaningful manner. How these two concerns were addressed is summarized in the following description of the project components and its results.

1. Organization of the Planning Project

The Assessment Planning Project aimed for broad input from groups with an interest in a state-by-state assessment in order to build a consensus of support among states, educators, and policy-makers. There were four important elements of the Project organization in developing this consensus:

- o **Steering Committee** -- representatives of 18 education organizations, including teacher unions, administrators, state organizations, mathematics educators, and education researchers -- made policy recommendations and established the agenda and structure for the Project;

- o **Mathematics Objectives Committee** -- a technical working committee comprised of mathematicians, state math supervisors, teachers, math assessment specialists, and a school principal -- prepared a draft report on mathematics objectives and a framework for assessment design;

- o **Analysis and Reporting Committee** -- a technical working committee comprised of state assessment directors, education researchers, statisticians, and a local district test director -- prepared a draft report on how to categorize and compare states' assessment results and identified demographic, policy, and program data needed to analyze the results.

- o **State committees of math teachers and curriculum specialists** -- in each state, the draft reports were reviewed and commented on for use in revisions by the technical committees.

2. Purposes of State Comparisons

The Steering Committee established recommendations on the purpose of state comparisons and the conditions that should be met for making such comparisons:

> The purpose of State Level Student Achievement Comparison is to provide data on student performance to assist policy-makers and educators to work toward the improvement of education. Such data can be useful by encouraging and contributing to a discussion of the quality of education and the conditions that determine it (CCSSO, 1988b).

Three specific contributions that state comparative assessments which are appropriately designed can make to education are in identifying problems in the curriculum or student learning, determining which programs improve performance, and measuring progress over time (Cody, 1987). Examples of the uses of state comparative results in these areas are:

- o **Identify and select problems**, such as needed improvement efforts with:
  - Subjects or topics in the curriculum,
  - Racial or ethnic groups in the student population,
  - Different levels of performance, e.g., basic vs. advanced,
  - Students in a region or by urban, suburban, rural.
- o **Determine programs and practices that make a difference**, such as:

  - Resource investments, e.g., per pupil expenditures, calculators, computers, textbooks
  - Teacher training, inservice, and support programs
  - Policies on instructional time and course requirements, as reflected in data on students opportunity to learn the topics assessed
  - School effectiveness programs

- **Measure progress, such as capacity for:**
  - Consistent, reliable trend data from a representative sample of students in the state
  - Analysis of presence or absence and degree of improvement in student learning
  - Providing feedback related to state reform efforts.

3. Assessment in Mathematics

The Mathematics Objectives Committee (CCSSO, 1988b) carried out three essential tasks: first, the Committee established a set of principles upon which a mathematics assessment should be based; second, the Committee designed a framework for the assessment which specifies mathematics abilities and content areas; and third, the Committee wrote sample items which illustrated how each type of objective should be assessed.

Three principles for the state-by-state mathematics assessment were specified:

a. Be inclusive of state curricula and mathematics knowledge, not "lowest common denominator," but rather dynamic and forward-looking.

The committee decided the assessment needed to incorporate the states curriculum objectives, but that it could not be simply a consensus of what is common across all the states current objectives. The assessment should look forward as to where curriculum development, instruction, and student learning should be going and it should be inclusive of different levels of mathematics knowledge (basic and advanced) as well as different approaches to teaching. The three key sources for the math objectives reflect this principle:

- Curriculum and Evaluation Standards for School Mathematics, final draft report of the National Council of Teachers of Mathematics;
- States curriculum frameworks, standards, and objectives, and review by the states of the Committee's work;
- Objectives for the 1986 NAEP Mathematics Assessment, to link the 1990 assessment to previous assessments.

b. Develop a relatively simple framework for assessment objectives to facilitate interpretation and understanding of test results.

The 1986 NAEP math objectives were organized in seven content and five process categories producing a matrix of 35 cells. The Committee favored fewer categories, recommending that desired complexity in areas of process and content be attained through specific topics in content areas. The recommended framework, or matrix, is displayed in Figure 5. The matrix arrays three areas of mathematical abilities (conceptual understanding, procedural knowledge, problem solving) by five content areas (numbers and operations, measurement, geometry, data analysis, statistics and probability, and algebra and functions).

c. In designing and testing items based on the mathematics objectives, do not treat content and mathematical ability categories as discrete or mutually exclusive categories.

Abilities and content areas are typically integrated and overlapping. Thus, some assessment exercises need to be multi-dimensional and designed to cut across content categories. The Committee recommended the weight that should be given to each category in the distribution of exercises by grade level of the assessment, as shown in Figure 6. The distribution is not specified according to a matrix in order to emphasize the need for integrating abilities and content. Conceptual understanding exercises, for example, are to be distributed across content areas.

4. Analysis and Reporting of Assessment Results State-by-State

The recommendations of the Analysis and Reporting Committee (CCSSO, 1988c) focused on two important principles for attaining fair and valid state

Figure 5

MATHEMATICS ASSESSMENT

THE FRAMEWORK

The mathematics assessment framework is organized by mathematical abilities and content areas. The mathematical abilities assessed are:

Conceptual Understanding
Procedural Knowledge
Problem Solving

Content is primarily from elementary and secondary school mathematics up to but not including calculus. The mathematics content areas assessed are:

Numbers and Operations
Measurement
Geometry
Data Analysis, Statistics, and Probability
Algebra and Functions

Framework for the Fifth Assessment

CONTENT AREAS

MATHEMATICAL ABILITIES	Numbers & Operations	Measurement	Geometry	Data Analysis Statistics & Probability	Algebra & Functions
Conceptual Understanding					
Procedural Knowledge					
Problem Solving					

Figure 6

TABLE 1: PERCENTAGE DISTRIBUTION OF EXERCISES BY GRADE AND MATHEMATICAL ABILITIES

Mathematical Abilities	Grade 4	Grade 8	Grade 12
Conceptual Understanding	40	40	40
Procedural Knowledge	30	30	30
Problem Solving	30	30	30

TABLE 2: PERCENTAGE DISTRIBUTION OF EXERCISES BY GRADE AND CONTENT AREA

Content	Grade 4	Grade 8	Grade 12
A. Numbers and Operations	45	30	10
B. Measurement	20	15	10
C. Geometry	15	20	25
D. Data Analysis, Statistics & Probability	10	15	20
E. Algebra & Functions	10	20	35

comparisons with assessment data.

a. **Test scores should not be reported by simply an average score of students assessed in each state; far more information is needed.**

b. **State comparative assessments need to include demographic data on each state and its students, as well as measures of state education policies, programs, and practices that may be related to student assessment scores.**

The recommended methods for reporting state scores are illustrated in examples of reporting formats in Figures 7 and 8. First, mathematics results should be reported according to a scale, similar to NAEP's proficiency scales, to show the percent of students in each each grade who score in or above each of the defined proficiency levels (e.g., in 1986, a level of 150 means accomplishment of "simple arithmetic facts," while 350 means accomplishment of "multi-step problem-solving and algebra."). Results should be reported according to the sub-scale scores for content and ability areas. Second, the state scores should be broken out by gender, race/ethnicity, and type of community, as well as showing trends over time (e.g., '90 to '92 assessments). Figure 7 shows the recommendation for reporting math achievement scale scores over time and by gender. Figure 8 shows the recommendation for reporting sub-scores for math achievement over time and by race/ethnicity.

The recommendations on state-by-state comparisons included use of state demographic variables, such as per capita income, percent of adults with 4 years high school, and percent of school age population in poverty, with rank-ordering of states by a composite index of these variables. Figure 9 shows an example of state rankings with such an index and reporting scale scores with the number of students scoring at each level and an average score for the state. The Committee did not support the idea of using either expected scores (vs. actual scores) or "adjusted scores" through regression statistical analysis with demographic variables; instead recommending that state scores be reported next to the state

Figure 7

State Mathematics Proficiency For All Students And By Gender

(State)　　　　Grade 4

	All Students			Female			Male		
Scale	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
150	___	___	___	___	___	___	___	___	___
200	___	___	___	___	___	___	___	___	___
250	___	___	___	___	___	___	___	___	___
300	___	___	___	___	___	___	___	___	___
350	___	___	___	___	___	___	___	___	___
Average	___	___	___	___	___	___	___	___	___

GRADE 8
(The Same)

GRADE 12
(The Same)

Figure 8

State Mean Sub-Scale Score By Type Of Mathematics Question By Race/Ethnicity

(State) Grade 4

Ability	Black Students			Hispanic Students			White Students		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Conceptual Understanding	___	___	___	___	___	___	___	___	___
Procedural Knowledge	___	___	___	___	___	___	___	___	___
Problem Solving	___	___	___	___	___	___	___	___	___
Combined (# Items)	___	___	___	___	___	___	___	___	___

Topic	Black Students			Hispanic Students			White Students		
	'90	'92	Chg.	'90	'92	Chg.	'90	'92	Chg.
Numbers and Operations	___	___	___	___	___	___	___	___	___
Measurement	___	___	___	___	___	___	___	___	___
Geometry	___	___	___	___	___	___	___	___	___
Data Analysis Statistics & Probability	___	___	___	___	___	___	___	___	___
Algebra & Functions	___	___	___	___	___	___	___	___	___
Combined (# Items)	___	___	___	___	___	___	___	___	___

GRADE 8
(The Same)

GRADE 12
(The Same)

Figure 9

STATES RANKED ON COMPOSITE INDEX OF DEMOGRAPHIC VARIABLES

STATES	1986 Per Capita Income	1980 % Adults 4 yrs H.S.	1980 % School Age Pop. in Poverty	Achievement Grade 4 Proficiency Scales 150	200	250	300	350	Aver.
Alaska	17,796	82.5	11.0	___	___	___	___	___	___
Nevada	15,437	75.5	9.0	___	___	___	___	___	___
Connecticut	19,600	70.3	10.2	___	___	___	___	___	___
New Hampshire	15,911	72.3	8.7	___	___	___	___	___	___
Wyoming	12,781	77.9	7.4	___	___	___	___	___	___
Colorado	15,234	78.6	10.5	___	___	___	___	___	___
Washington	15,009	77.6	10.0	___	___	___	___	___	___
Minnesota	14,994	73.1	9.3	___	___	___	___	___	___
Massachusetts	17,772	72.2	12.1	___	___	___	___	___	___
Hawaii	14,886	73.8	11.4	___	___	___	___	___	___
Kansas	14,650	73.3	10.5	___	___	___	___	___	___
Oregon	13,328	75.6	10.4	___	___	___	___	___	___
California	16,904	73.5	13.8	___	___	___	___	___	___
New Jersey	18,626	67.4	13.2	___	___	___	___	___	___
Wisconsin	13,909	69.6	9.5	___	___	___	___	___	___
Maryland	16,864	67.4	11.6	___	___	___	___	___	___
Utah	10,981	80.0	9.6	___	___	___	___	___	___
Nebraska	13,742	73.4	11.4	___	___	___	___	___	___
Iowa	13,348	71.5	10.6	___	___	___	___	___	___
Michigan	14,775	68.0	12.2	___	___	___	___	___	___
Illinois	15,586	66.5	13.9	___	___	___	___	___	___
Delaware	15,010	68.6	14.4	___	___	___	___	___	___
Montana	11,803	74.4	12.5	___	___	___	___	___	___
Ohio	13,933	67.0	12.0	___	___	___	___	___	___
Vermont	13,348	71.0	12.7	___	___	___	___	___	___
New York	17,111	66.3	17.5	___	___	___	___	___	___
Indiana	13,136	66.4	10.8	___	___	___	___	___	___
Arizona	13,474	72.4	15.4	___	___	___	___	___	___
Virginia	15,408	62.4	14.1	___	___	___	___	___	___
Pennsylvania	14,249	64.7	13.0	___	___	___	___	___	___
Rhode Island	14,579	61.1	12.4	___	___	___	___	___	___
Idaho	11,223	73.7	13.1	___	___	___	___	___	___
Florida	14,646	66.7	17.2	___	___	___	___	___	___
Missouri	13,789	63.5	13.7	___	___	___	___	___	___
Maine	12,790	68.7	14.8	___	___	___	___	___	___
North Dakota	12,472	66.4	13.7	___	___	___	___	___	___
Oklahoma	12,283	66.0	14.7	___	___	___	___	___	___
Texas	13,478	62.6	18.1	___	___	___	___	___	___
South Dakota	11,814	67.9	19.0	___	___	___	___	___	___
New Mexico	11,422	68.9	21.2	___	___	___	___	___	___
Georgia	13,446	56.4	20.1	___	___	___	___	___	___
North Carolina	12,438	54.8	17.5	___	___	___	___	___	___

rankings on demographic characteristics. A second recommendation was to compare state achievement scores by disaggregating scores according to student socio-economic background. For example, even though the number of economically disadvantaged students to be tested in each state would vary, scale scores could be reported for the group of disadvantaged students in the state sample to determine differences in achievement for this sub-population.

Data should be collected on measures of school processes, teacher characteristics, and instructional practices that have been shown to be related to student achievement. It was recommended that these variables be limited to those that can be affected by state policies, programs, and practices, and by what is reasonable to ask of students, principals, and teachers and by the quality of data that can be obtained. Other surveys and research studies may be more appropriate for answering broader research questions about educational organization and processes. The specific types of education variables recommended for the state-by-state national assessment are listed in Figure 10.

SUMMARY

This paper has outlined the approach of the Council of Chief State School Officers to developing and implementing a model for comparative state education indicators. The process for putting these indicators in place has emphasized a high degree of involvement by state education specialists and by experts in education research, measurement, and statistics. A set of standards for selecting education indicators has been consistently applied to ensure that state comparisons are valid and meaningful. The descriptions of two projects demonstrate that the process of developing state-by-state indicators differs according to the source of data for indicators. However, the descriptions also show similarities in the use of analysis of what information is important to

Figure 10

EDUCATION VARIABLES

- The number of mathematics courses the student has completed in high school. (Grade 12)
- The types of mathematics courses the student has completed or the advanced level courses the student has completed such as Algebra 2, Geometry, Statistics. (Grade 12)
- The amount of time the student spends on mathematics homework. (Grades 4, 8 & 12)
- The actual time (minutes per week) that students receive mathematics instruction. (Grades 4, 8 & 12)
- Whether students have had the opportunity to learn through class instruction specific topics within a process or content area of mathematics. Have certain topics been part of the scheduled and presented instructional program? (Grades 4, 8 & 12)

- Whether teachers of mathematics have a state certificate to teach mathematics. (Grades 4, 8 & 12)
- Years of experience teaching mathematics. (Grades 4, 8 & 12)
- Number of courses or credits teachers of mathematics classes received in college and graduate school in mathematics and in mathematics pedagogy. (Grades 4 & 8)
- Whether teachers of mathematics have college majors or minors in mathematics. (Grades 4, 8 & 12)
- The extent of the mathematics teachers' recent professional development in mathematics or mathematics education (graduate courses, inservice workshops, conferences, meetings, etc.). (Grades 4, 8 & 12)
- Per pupil expenditures.

The specific method of collecting data on the above variables is beyond the purview of this particular report. NAEP has access to technical experts on its staff and elsewhere to design the appropriate questions or methods to secure useful data.

report about education progress in specific subject areas, involvement of key groups of producers and users of indicators, and careful consideration of how information should be reported to maximize the uses of the indicators.

REFERENCES

Blank, Rolf. (1986). "Conceptual Framework for Developing a State-Based Network for Science and Mathematics Indicators," Washington, D.C.: Council of Chief State School Officers.

Blank, Rolf and Espenshade, Pamela. (1988). "State Education Indicators on Science and Mathematics: Results of a State-by-State Inventory." Washington, D.C.: Council of Chief State School Officers.

Blank, Rolf and Selden, Ramsay. (1989). "Alternative Methods of Assessing Science: Report to the States." Washington, D.C.: Council of Chief State School Officers

Cody, Wilmer S. (1987). "Developing State Level Report Cards of Student Achievement. Washington, D.C.: CCSSO.

Council of Chief State School Officers. (1985). "Assessment and Evaluation Plan of the Committee on Coordinating Education Information and Research." Washington, D.C.: CCSSO.

Council of Chief State School Officers. (1988a). State Education Indicators 1988. Washington, D.C.: CCSSO.

Council of Chief State School Officers. (1988b) "Assessing Mathematics in 1990 by the National Assessment of Educational Progress." National Assessment Planning Project. Washington, D.C.: CCSSO.

Council of Chief State School Officers. (1988c) "On Reporting Achievement at the State Level by the National Assessment of Educational Progress." National Assessment Planning Project. Washington, D.C.: CCSSO.

Council of Chief State School Officers/State Education Assessment Center. (1988). "Instructions and Reporting Forms for Data on Science and Mathematics Education in (each state)." Washington, D.C.: CCSSO.

Murnane, Richard J. & Senta A. Raizen (eds.) (1988). Improving Indicators of the Quality of Science and Mathematics Education in Grades K-12. National Research Council. Washington, D.C.: National Academy Press.

National Center for Education Statistics. (1987). The Condition of Education. Washington, D.C.: U.S. Department of Education.

National Governors' Association. (1986). Time for Results: The Governors' 1991 Report on Education, Washington, D.C.: National Governors' Association, Center for Policy Research and Analysis.

National Science Board. (1987). Science Indicators: The 1987 Report. Washington, D.C.: National Science Foundation.

Oakes, Jeannie. (1986). Educational Indicators: A Guide for Policymakers. New Brunswick, N.J.: Center for Policy Research in Education.

Phi Delta Kappan. (1988). Six articles on Education Indicators, Phi Delta Kappan, March, pp. 486-513.

Shavelson, Richard, Lorraine McDonnell, Jeannie Oakes, & Neil Carey. Indicator Systems for Monitoring Mathematics and Science Education. Santa Monica, CA: RAND Corporation, 1987.

U.S. Department of Education, Office of Educational Research and Improvement. (1988). Creating Responsible and Responsive Accountability Systems. Washington, D.C.: U.S. Department of Education.

RESEARCH, NATIONAL ASSESSMENT, AND SCHOOL EVALUATION IN THE U.K.

Caroline Gipps
Institute of Education (United Kingdom)

National Assessment and School Evaluation in the UK

This paper addresses the inter-relationship between national assessment and school evaluation in the UK. There are two forms of national assessment in the UK: the work carried out by the Assessment of Performance Unit (APU), the British equivalent of NAEP, and the national assessment programme which from 1991 will assess the new national curriculum.

1. National Assessment

The Assessment of Performance Unit

The APU is a unit within the Government Department of Education and Science (DES) which supervises the national assessment of performance in maths, language, science, modern languages and design for technology. Although the APU was set up in 1975 at a time of concern over the education of minority children, and has as one of its tasks to identify 'underachievement', in reality its main task, as far as the DES was concerned was to operate as an indicator of educational 'standards' and to give ministers information on whether, and by how much, these were rising or falling.

In the growing atmosphere of accountability in the late 1970s, when it became clear that the APU was intending to monitor standards, there was considerable concern that the APU was intended as an instrument to force accountability on schools and therefore teachers. The APU assessment programme, though ostensibly concerned with children's standards, was interpreted as potentially dealing with teachers' competencies. The teaching unions, therefore, viewed it with great concern. Our 1982 evaluation (Gipps & Goldstein, 1983) however, showed that this fear did not materialise. The idea of accountability of teachers related to pupil assessment came as a considerable shock to British educators and the teacher unions were allowed to bargain for strict anonymity and confidentiality of data, so that test results could not be used to evaluate individual schools. In the event, the decisions to go for light sampling, matrix sampling and anonymity of pupils and schools (the latter being a firm requirement of teaching unions), precluded any role in teacher and school evaluation.

The APU has also had only limited success in monitoring changes in performance over time due to a major technical problem. That is, changes large enough to be meaningful will only be detected over a number of years, at least four or five, and any serious monitoring of performance would go on over a longer period than that. For example national Reading Surveys ran from 1948 - 72. The problem is that the same test used over this sort of period becomes dated. Curriculum and teaching change, thus the test becomes harder and standards will seem to fall. To make the test 'fair' it is necessary to update it, but then one cannot compare the results on the modified version of the test with the results on the original form because it is not a true comparison.

> "The problem here is that various statistical techniques are needed to calculate comparable difficulty levels and there is no consensus on which of them is satisfactory. It seems that 'absolute' measures of change over time are difficult if not impossible to obtain and, at most, we can only hope to measure 'relative' - between group - changes." (Goldstein, 1983)

In the early 1980s the APU had to drop the controversial Rasch technique of analysing difficulty levels of test items and admit that it could not comment on trends over time. This issue has dogged the APU's work for several years (see Nuttall, 1986), and is by no means a uniquely British problem. NAEP has also grappled with it and relies on using 'item response models' which are elaborated versions of the Rasch technique.

With the problems of measuring changes in performance over time, the APU's aim of monitoring standards, and thus the performance of schools in general, became weakened. The test development teams, with their background in subject-area research rather than psychometrics, sought to resolve this situation by using the data for research purposes.

Each of the three major survey areas, maths, language and science, completed an initial round of five annual surveys in 1982, 1983 and 1984 respectively. After this initial phase, the teams were commissioned to survey only every five years and to spend time in between surveys on dissemination and on making a more detailed analysis of their findings, for example, in relation to school and child background factors. This was referred to as 'mining the data'.

There is no doubt that the findings from the test development teams include a tremendous amount of information of use to teachers, whether it is about children's errors in maths, children's misconceptions in science or the linking of reading, writing and cracy skills in language. The APU is committed to using the implications of the survey findings to help teachers, and clearly, it would be a waste not to do so.

It is ironic that this national assessment, which was greeted with such fear and concern by many in the teaching profession, became a research exercise with direct and valuable feedback to practitioners (Gipps 1987).

National Curriculum Assessment

The proposed programme of national assessment to monitor the national curriculum is however, a very different exercise. All children of 7, 11, 14 (and 16 for subjects not assessed via the school leaving exams) will be assessed on tests and activities directly related to the national curriculum. The proposals, from the Task Group on Assessment and Testing (TGAT), are complex (DES 1988) but, in summary, the arrangements are that each of the core and foundation subjects is to be divided up by subject working groups into a number of components (eg listening, speaking, reading and writing for language). Attainment targets, which are descriptions of knowledge and activities to be learned, (that is what children should 'know, understand and be able to do') cluster within the components. Each of these attainment targets is divided up into 10 levels of performance covering the age range 7 to 16. Children are to be assessed on these attainment targets at these levels by a mixture of testing and teacher assessment and this will give profiles of attainment.

There will be supplementary assessment available for any children who need further investigation; these will be largely diagnostic. For children with special educational needs, who are deemed not able to take the national tests, there will eventually be test materials and assessment procedures specially designed for them.

These profiles of attainment are to serve a formative assessment function at 7, 11 and 14 - i.e. to guide the child's future teaching and learning programme. They will also be used as a basis for communication with parents in records of achievement. This detailed, structured information will no doubt be very valuable.

But this detailed information will also, however, be summarised for publication. Children will be assigned to a level of attainment in each subject from one to ten, each level representing two years' progress. Level two will be the expected seven year old performance level, level four the expected eleven year old performance level etc. Schools will have to publish distributions of performance on these levels at 11, 14 and 16. Although the Task group does not recommend publication at 7, Mr. Baker, Secretary of State for Education and Science does, so it is likely that 7 year old results will also be published.

Not only must results be made publicly available for each school, but to parents and those responsible for the school for the relevant classes also. Results for the class may well emerge in the local area as measures of teacher effectiveness. On the basis of the test scores, local league tables can be formed and parents will make their choice of school.

These school level results are *not* to be adjusted for socio-economic background of the intake. The TGAT Report argues that using statistically adjusted results to compare schools' performance 'would be liable to lead to complacency if results were adjusted and to misinterpretation if they were not'.

Instead, for each school, the results will be set in the context of a written account of the work of the school as a whole, and socio-economic and other influences that are known to affect attainment. There is a considerable amount of research on adjusting exam scores for intake and on measuring school effectiveness which TGAT simply ignores (Nuttall, 1988). As Goldstein and Cuttance (1988) point out, the attainment of the children entering a school is the single most important determinant of subsequent achievement. There is a large literature showing that children from socially disadvantaged areas tend to have lower exam scores and test results than those from more socially advantaged areas (e.g. Coleman et al 1966; Plowden 1966, Rutter et al 1979). Simple school test and exam averages will in part reflect these differences and the students' performance on intake, and thus obscure any real "effects" due to the school.

The solution of TGAT : to publish results aggregated for the school in the context of a general report for the area ... "to indicate the nature of socio-economic and other influences which are known to affect schools", suggests that allowances may be made by parents and others, but at a much less precise level than that of statistical adjustment. Indeed, the responsibility for the interpretation is passed on to the audience, that is to parents and the general public.

The reason that statistical adjustment is not to be used is that if the comparison is a direct one parents can look for schools with actual high scores - which is what, the Government says, most parents want. This is of course partly true, but what this government ignores is that sophisticated analysis can tell us which schools are performing well for particular groups of pupils, i.e. girls/boys, ethnic minorities, etc., in which parents will most certainly be interested.

As the assessments work their way through the system there will be 11 year old data against which to analyse the 14 and 16 year old scores, and similarly 7 year old data against which to analyse 11 year old scores; but for many schools damage to their image and popularity will have been done by then. There is of course scope for an individual Local Education Authority (LEA) i.e. School District, to choose to adjust its test results for intake, but how many LEAs will find the facilities to do this if it is not a requirement?

The aggregated assessment results are to be made available in a standard format, this is apparently not so that LEAs can publish league tables, but so that each school's results can be set against those for the LEA as a whole. And eventually, LEAs themselves will be compared and evaluated: "In due course, LEAs should also be required to submit to the Secretary of State data on distributions of attainment at the four key ages with comparisons over time for all schools they maintain, as the basis for compiling national data and so that the SEAC and the NCC can monitor standards of attainment..." (DES, 1989)

Although many teachers and their unions are unhappy with aspects of the national curriculum and assessment proposals they have been powerless to resist their implementation or indeed to modify the proposals. What is more surprising is that subject working groups, containing hand picked subject experts have had their proposals for the national curriculum quite drastically modified at DES level.

2. School Evaluation

How then have schools been accounting for themselves over the last fifteen years, since APU national assessment has not been able to deliver the monitoring of school performance ?

School Self-Evaluation

School self-evaluation (SSE) has been a major development. This trend began in the mid-seventies and the emphasis was on critical examination by schools of their organisation, processes and/or outcomes. As Simons (1988a) points out, SSE had several purposes : as a response to accountability demands, certainly, but also for professional development and to improve managerial efficiency. The process also took a variety of forms: organisational analysis, process oriented or issue based approaches and checklists for self-review. Despite these various purposes and approaches, in general the focus was, and is, on the institution. The institution, not the individual, is the unit of evaluation and change : teacher development and pupil performance issues are considered in the context of whole school policies.

This movement, with its emphasis on a professional body carrying out evaluation in its own terms (and research too within an action research framework) to improve professional practice and understanding, was in reaction against managerial and productivity models of accountability, based on economic and administrative concerns for efficiency, tidiness and value for money (MacIntyre, 1989).

A major issue for SSE in an accountability setting is that, although the outcomes of SSE may be made public, they are rarely read by the public. Indeed, in some LEAs there is concern that not even LEA staff read the (lengthy) SSE reports. For schools, however, the importance lies in the process, not the product. SSE is, therefore, largely an 'invisible' method of evaluating schools and is not generally in the public domain. Thus, by the early 1980s it became clear that professional, school based self-evaluation was not acceptable to politicians (and many parents) as a basis for school evaluation in the era of 'value for money' and market place competition.

Publication of Exam Results

The first manifestation of this came in the 1980 Education Act, when secondary schools were required to publish exam results. This Act was the first of the new Conservative Government's moves in education and a major plank of the Act was that parents using the state sector should have more information and choice in deciding which schools their children should attend. Thus, from 1982 the annual publication of public examination results (at 16 & 18) was made compulsory.

Reaction to this requirement to publish examination results was mixed: several of the teachers' unions objected on professional grounds, believing that the published information was likely to be misleading and to have a deleterious effect on the education provided by secondary schools; local authority organisations argued that the expense of providing the information could not be justified. On the other hand, there was a widespread belief that schools should be more accountable to the communities they serve and that the publication of examination results would help to bring this about.

There is no compulsion for schools to provide summary measures of these results such as total number of passes and pass rates, although some may choose to do so. The regulations were designed to make school 'league tables ' difficult to construct, but certainly in London, the national evening newspaper has published league tables of London's secondary schools.

Since schools with intakes of mainly below average pupils at eleven will not usually get and, indeed, cannot be expected to get, for reasons explained earlier, examination results which are as good as schools which get a large proportion of above average pupils at eleven, researchers and LEAs began developing ways of analysing school examination results taking into account 'background variables' (see Plewis et al 1981). This is how the work referred to in the section (above) on publishing National Curriculum Assessment results was begun.

The publication of secondary school exam results has not had as much of an effect on schools as perhaps might have been anticipated. Parents had another piece of information on which to base their choice of school, but schools have not become obviously more efficient, nor has there been widespread closing down of schools with poor exam results.

Part of the function of the 1988 Education Reform Act is to remedy this situation. It strengthens parental choice, widens the net of publication of (national assessment) results to include junior schools, and strengthens the market place model with regard to the fate of 'popular' and 'unpopular' schools.

Performance Indicators

There is, also, as a separate development, the emergence of performance indicators which are being piloted in eight LEAs. The idea is that schools' performance and 'value for money' are to be calculated on the basis of a very wide range of factors including : pupil teacher ratios, qualifications of staff, class management and teaching skills, teachers' commitment and professional attitudes, the quality of curricular management, management of time; students' engagement in the learning process, quality of the learning experience, outcomes of learning, homework policy, attendance and punctuality, attitudes and behaviour including incidents of vandalism and graffiti; and costs per pupil (SIS 1988). Of course, at a crude level, school performance indicators are nothing new. Exam results, sports results, the annual school concert have always been used by the public to evaluate schools. However, the increasing emphasis on 'value for money', accountability and parental choice has resulted in attempts to develop more 'scientific' approaches to the assessment of schools' performance.

It is clear, however, that developing sophisticated performance indicators which incorporate qualitative as well as quantitative measures will be a very complex task. At the moment there does seem to be a commitment to including qualitative information, but some indicators will not lend themselves particularly well to quantification, and we know that where we have numerical and descriptive information, the quantitative data tends to overwhelm the qualitative information.

Performance indicators are still in the developmental stage, but there is no doubt that the 'economics' model of evaluating schools is with us for the foreseeable future. Whether complex performance indicators are used or whether simple league tables of schools on the basis of published national assessment results prevail, is hard to predict at this moment.

My colleague Helen Simons has said that in schools where the professional self-monitoring role is well-established it may be that teachers are so reluctant to give up this method of professional development that they continue with it, somehow fitting it in with the other evaluation and national curriculum demands. Simons even goes so far as to suggest that, rather than schools competing against each other, which is an intended outcome of the 1988 Education Reform Act, (see Gipps 1988a) that local groups of schools will collaborate for purposes of pooling resources, e.g. support services , offering mutual support through a period of change, and supporting professional development (Simons, 1988b). Again, we shall have to wait and see.

3. Discussion

I want to focus this discussion around three questions:

Will national curriculum assessment raise standards?
Will the national curriculum assessments be adequate as measures of school evaluation?
What do these developments mean for teachers' professional status and power ?

There is little evidence that the introduction of mandated testing raises 'standards' short of teaching to the test (Gipps 1988b, BJES). However, what we have in the UK is the introduction of mandated testing linked to specific curriculum objectives (the attainment targets and statements of attainment) and a high significance placed on the results. This combination is unprecedented on a national scale and will, I suspect, be a particularly powerful tool for controlling and evaluating schools, teachers and pupils.*

---

* Broadfoot, too (1988) has outlined the strong surveillance potential inherent in the national curriculum, assessment and recording arrangements.

There is little doubt about the significance of national curriculum assessment results. Students are to be graded and classified; schools, and indeed some teachers, are to be evaluated on the basis of the results; if they are found wanting, schools are expected to close and teachers may face redundancy, or lack of promotion. That is pretty significant. Even if a particular teacher's job is not on the line, nor the school unpopular, teachers will be conscious of the importance of the assessment results for childrens' life chances. The significance, therefore, will operate at every level.

The perception of significance is, and always has been, crucial to the impact of assessment. Our primary schools are currently full of standardised tests which are seen as routine and useful rather than significant, and they have little impact on teaching or curriculum (Gipps et al 1983). Twenty five years ago, however, primary schools were dominated by the demands of a highly significant selection exam at 11+.

This is what is called in the American literature 'high stakes' testing. Madaus (1988) has analysed the relationship between the level of stakes a test is perceived to have and the effects of the test, and posited seven principles :

- The power of tests and examinations to affect individuals, institutions, curriculum or instruction is a perceptual phenomenon : if students, teachers, or administrators believe that the results of an examination are important, it matters very little whether this is really true or false - the effect is produced by what individuals perceive to be the case;

- the more any quantitative social indicator is used for social decision-making, the more likely it will be to distort and corrupt the social processes it is intended to monitor;

- if important decisions are presumed to be related to test results, then teachers will teach to the test;

- in every setting where a high-stakes test operates, a tradition of past exams develops, which eventually de facto defines the curriculum;

- teachers pay particular attention to the form of questions on a high-stakes test, (e.g., short answer, essay, multiple choice), and adjust their instruction accordingly;

- when test results are the sole or even partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement;

* a high-stakes test transfers control over the curriculum to the agency which sets or controls the exam.

Corbett and Wilson (1988), describing the effects of raising the stakes in two state minimum competency testing programmes, found support for these principles, in particular that the aim becomes to improve the next set of test scores rather than some longer-term, more general goal of improving student learning. Regardless of the teachers' personal and professional opinions about the tests the fact was that students had to pass and teachers felt responsible to ensure they did. It would take a very unprofessional teacher to ignore the demands of a life-controlling assessment for his or her students.

But what about the wider professional implications for teachers? In Maryland, Corbett and Wilson (op. cit.) found that teachers felt that the mandated MCT programme had narrowed and improved the curriculum in terms of both course objectives and the range of courses offered. This was as a consequence of aligning the curriculum with test objectives. Our national assessment, though not an MCT programme, is of course to be linked to a national curriculum and there is considerable concern that this will have serious implications for teachers' professional role with regard to curriculum choice and development. This quote encapsulates a widely held view among British researchers and teacher educators :

> "The involvement of teachers in a generative role incurriculum determination is essential to good practice. Teachers make poor operatives of other people's ideas." (Simons, 1988, p.86).

Well, this may be a particularly British point of view because as Broadfoot (1988, op. cit.) points out French and German teachers are able to live with it. French teachers are apparently strongly conservative with no enthusiasm for, or experience of, teacher-led development work. An English report of the West German education system (DES, 1986) indicates that German teachers too, concentrate on delivery and assessment rather than curricular development.

How our teachers will react to the full range of curricular changes is impossible to predict, but I suspect it will not be very different from the Maryland teachers as described by Corbett and Wilson:

> "A clearer definition of what was expected to be covered represented an improvement over rambling curriculum guides but at the same time did exclude some content that staff members previously had deemed worthy of inclusion. Up to a

> point, Maryland educators viewed a tighter curriculum as a better one; they worried, however, that the trend would lead to excessively basic course offerings." (op. cit. p.30)

Whatever the future holds for teacher professionalism, it is salutary to consider how we have reached this state of affairs. Why is it that teachers, and the educational establishment, have been powerless to affect the path of current developments, when by contrast they won significant victories over the way the APU was run?

The answer lies partly in the role of the teaching unions. The crucial APU committee, the Consultative Committee, which made key recommendations to the DES (all of which were accepted) contained a large teacher union block, which tended to control the direction of voting (Gipps and Goldstein, 1983).

Today, the presence of even a single teaching union representative on an advisory committee would be unlikely.

Over the last five years a period of industrial action, coupled with genuine disagreement about educational reform, has weakened the unions. This has culminated in their failure to unite in the face of the 1988 Education Reform Act.

But, to be fair to the unions, this has to be set against a more general erosion of teachers' rights. The 1987 Teachers' Pay and Conditions Act resulted in the loss of negotiating rights over salaries and conditions of employment. It established a clear contract of responsibilities, directed how teachers' time was to spent and required appraisal procedures to be developed. The 1988 Education Reform Act completes the process by removing autonomy over curriculum, assessment and reporting. The shift in power is towards the Secretary of State, although the moves towards greater choice for parents and greater power for school governors makes it look as though individuals are empowered. (Deem, 1988)

The most powerful indicator of the way things have changed is that schools which wish to engage in curricular innovation will in future have to apply to the Secretary of State for permission to do so. None of the teachers, I know are particularly optimistic about being given such permission. So much for institutional curriculumm development and the teacher as researcher.

But there is a tension for the Government inherent in the national curriculum developments. Force of law aside, they do need some co-operation and even goodwill from teachers. Thus teachers are promised that they can choose the pedagogy and plan the detailed schemes of work they use to deliver the national curriculum. The contradictions inherent in the suggestions for teachers' role in national assessment are a clear illustration of the tensions in the Government's position: they want to control the teaching force and yet they need their compliance and co-operation, so they must be seen to be offering something. The national assessments hinge on the use of professional teacher assessment: teachers have been sold a vision of national assessment which relies on teacher judgement. But the reality is that, through moderation, teacher judgement will be subject to the dominance of 'objective' test scores. In the event what teachers are they offered may turn out to be illusory.

Now to the final issue - will national curriculum assessment help to evaluate schools ?

There is no doubt that the results will be an important factor in school evaluation (even if, unadjusted, they are misleading and could result in the unjustified victimisation of schools in socially disadvantaged communities while failing to locate the poorly performing schools in the socially advantaged communities, as Goldstein and Cuttance (op. cit) point out).

But as we all know, and the TGAT Report itself acknowledges, a school's performance can onlybe judged fairly by taking account of many aspects of its work. At the same time, however, the TGAT Report proposes setting up a highly significant assessment system with inevitable consequences : that one indicator of performance becomes the indicator, and then the goal itself. In this model, school evaluation becomes straightforward, even simple, but what it loses is educational validity.

References

Broadfoot,P (1988) 'The National Assessment Framework and Records of Achievement' in National Assessment and Testing: A Research Response BERA

Corbett, H.D. & Wilson, B. (1988) 'Raising the stakes in statewide mandatory minimum competency testing' in The 1988 Year Book of the Politics of Education Association - Journal of Education Policy Vol. 3, No. 5

Deem, R. (1988) 'The Great Education Reform Bill 1988 - some issues and implications' in Journal of Education Policy Vol. 3, No. 2

DES (1986) Education in the Federal Republic of Germany - Aspects of curriculum and assessment HMI

DES (1988) National Curriculum: Task Group on Assessment and Testing: A Report DES Welsh Office

DES (1989) Draft Circular - The Education (School Curriculum and Related Information) Regulations 1989

Gipps, C. (1987) 'The APU: from Trojan Horse to Angel of Light' in Curriculum Vol. 8, No. 1

Gipps, C. (1988 a) 'The TGAT Report: Trick or Treat?' Forum Vol. 31, No. 1

Gipps, C. (1988 b) 'Standards in Education and the Role of Testing' British Journal of Educational Studies Vol. 36, No. 1, pp 21-36

Gipps, C. & Goldstein, H. (1983) Monitoring Children: an evaluation of the Assessment of Performance Unit. Heinemann Educational Books

Gipps, C. et al (1983) Testing Children: Standardised Testing in Schools and LEAs Heinemann Educational Books

Goldstein, H. (1983) 'Measuring changes in educational attainment over time: problems and possibilities' Journal of Educational Measurement Vol. 20, No. 4 Winter

Goldstein,H. & Cuttance, P. (1988) 'National Assessment and School Comparisons' Journal of Education Policy Vol. 3, No. 2

MacIntyre, A. (1989) Evaluating Schools Open University (in press)

Madaus,G.(1988) 'The influence of testing on the curriculum', in L. Tanner (ed) Critical Issues in

Curriculum 87th Year Book of NSSE, Part 1 University of Chicago Press

Nuttall, D. (1986) 'Problems in the Measurement of Change' in Nuttall, D. (ed) Assessing Educational Achievement Falmer Press

Nuttall, D. (1988) National Assessment: Complacency or Misinterpretation? Public Lecture given on 2nd March 1988 at ULIE

Plewis, I. et al (1981) Publishing School Examination Results - A Discussion Bedford Way Papers 5 University of London Institute of Education

Plowden (1967) Children and their Primary Schools DES

Rutter et al (1979) Fifteen Thousand Hours Open Books

Simons, H. (1988 a) 'Teacher Professionalism and the National Curriculum' in The National Curriculum Bedford Way Papers No. 33

Simons, H. (1988 b) Evaluation and the Reform of Schools ULIE

Statistical Information Service (1988) Performance Indicators for Schools - a Consultation Document The Chartered Institute of Public Finance and Accountancy

THE RANKING OF SCHOOLS AND THE FLORIDA NORM-REFERENCED TESTING PROGRAM

Marc Baron
Palm Beach County Schools

Introduction

The Florida Norm-Referenced Testing Program (FNRTP) was developed by the Florida Department of Education (FDOE) as a consequence of the FACET Act. The program was first implemented during the 1986-87 school year. A test was produced by the FDOE that was administered to third grade students in Florida during the spring of 1987. In the fall of the 1987-88 school year test results were disseminated that ranked the schools and the districts in Florida. Issues associated with the ranking of schools and districts within Florida resurfaced upon the release of this information.

The emphasis on ranking in the FACET Act comes from the assumption that performance of employees at a school will improve when their school's relative standing or ranking is known to themselves and to the public. Is this assumption correct?

An Analogy

I believe an analogy may be in order here. Let us pretend that each of you had a child in third grade who took a reading comprehension test and that these test results were used to assess your effectiveness as parents by ranking each of you on the basis of your child's test scores.

What would your reaction be to such information? Think about your answer to this question. Would not some of you quickly, but appropriately, make the following comments and ask the following questions?

Let me see if I have this straight, my effectiveness as a parent is being determined by my child's ranking on a one hour test of reading comprehension. Is it reasonable to determine my rank (i.e., effectiveness) on the basis of this one test when there are many other indicators of my effectiveness as a parent?

In any event, although I might have had something to do with my child's reading comprehension skills, I have had a lot of help (high rankers) or hindrances (low rankers) along the way (e.g., my spouse and my child's grandparents). How can you determine my effectiveness as a parent when other factors, over which I have little or no control, contribute to my ranking based on my child's test score?

I hope the analogy makes clear that many of you would believe that the ranking you would have received would be unfair. It would be unfair because your ranking would be based on a test score that does not directly relate to your performance and that implies a valid overall ranking on the basis of this one test score. These and similar concerns and their relationship to the perceived fairness of using student test scores to assess employee performance relate to employee behavior. In the remainder of this paper I will provide both conceptual and practical information that demonstrates the following: (a) Student test scores are inappropriate measures of teacher and school performance. (b) practical issues make the generation of school rankings fallible, and (c) school personnel recognize the inappropriateness and fallibility of the rankings and will likely behave in ways inconsistent with the goals of a school.

A Conceptual Difficulty

Berk (in press), a national expert in the use of test scores, addressed the conceptual issue of using test scores to assess teacher and school performance. His response is as follows:

> Isn't student gain the most important product or work outcome of the job of teaching? After all, if a factory worker's productivity can be measured by the number of widgets he or she produces over a given period of time, why not evaluate a teacher's (or school's) effectiveness, that is productivity, by his or her students achievement gains at the end of the school year (cf. Medley, Coker, & Soar, 1984, p.33)?
>
> The arguments for this factory worker-teacher productivity analogy are derived from the principles of a piece-rate compensation system (Murnane & Cohen, 1986). Piece-rate contracts, where a worker is paid according to the number of widgets produced, is the most common form of "payment by results" (Pencavel, 1977). About 30% of the workers in U.S. factories are employed under piece-rate contracts (Seiler, 1984). These contracts provide a strong incentive for workers to produce because high productivity results in immediate rewards. When this piece-rate compensation system is applied to teachers, it breaks down because of the nature of the teaching process and the classroom environment.... Teachers must work with students whose individual characteristics vary considerably within a single class. This variability precludes all of the students from achieving the same amount at the same rate over 10 months.... Certain key characteristics of students, such as intelligence and home environment, markedly influence the quality and quantity of their academic achievement, irrespective of what the teacher does in the classroom. Consequently, a teacher's effectiveness is directly affected by the characteristics of the class, which is beyond a teacher's control.

Some Practical Dilemmas

Using One Score

Berk's comments have practical implications for policy makers. Fetler (1986), of the California Department of Education, in his summary of the current California school accountability program cautioned about using a single measure to rank schools and indicated that an assessment of a school's performance should take into account such factors as a school's "demographic and cultural context, curriculum and instructional methods, initial student achievement standing, subsequent student achievement in school...." Fetler further suggested that if school rankings are based on a single score they will be misleading because a single score cannot validly represent a school's effectiveness and does not consider how variable student achievement may be from school to school. For example, in Orange County Public Schools there were two schools with the same mean score (40), but the variability of student achievement at one school was 50% larger than at the other school. In words, although the school means were identical, since the scores of the students at one school were much more similar than they were at the other school, these schools were in fact very different in terms of the variability of the students' test scores. The FNRTP, however, generated the school ranks based only on school means, although the FNRTP did report the ranges and variability of student achievement for each school.

Accounting for Socioeconomic Differences

Fetler's comment about accounting for a school's "demographic and cultural context" reflects the fact that schools with high test scores tend to have students who come from homes with relatively high income levels and schools with low test scores tend to have students who come from homes with relatively low income levels. In other words, if a school's demographic and cultural context is not taken into account a school's rank will be largely determined by the socioeconomic composition of its students. If'schools with similar socioeconomic characteristics are compared, low socioeconomic schools can obtain high ranks and high socioeconomic schools can obtain low ranks. However, Fetler pointed out that by arbitrarily categorizing schools by socioeconomic status, schools near the category boundaries may be placed into a socioeconomic category as a result of the socioeconomic status of only a few children. As regards the FNRTP, Fetler's comments are applicable. Because the FNRTP has three socioeconomic categories, which are based on the percent of students receiving free or reduced lunch, the impression may exist that any differences among schools have been accounted for by the socioeconomic categorization. But the categories themselves were rather wide (e.g., Category A contained schools that had from 0% to 30%, Category B contained schools that had from 31% to 49%, and Category C contained schools that had from 50% to 100% of its students receiving free or reduced lunch). As a consequence, schools that had no students receiving free and reduced lunch were compared to schools that had one-third of its students receiving free and reduced lunch. Clearly, these schools do not belong in the same socioeconomic category. Thus, the socioeconomic categorization of schools used by the FNRTP to provide school rankings that accounted for demographic differences among schools was not entirely effective.

Determining School Ranks

The procedure used by the FNRTP to determine the ranks of schools and districts also had its associated difficulties. The scores received by the schools and the districts within Florida were similar enough so that it was necessary to develop ranks that were based on scores to the fourth decimal place. Ranks based on differences of .0001 of a point provided information that made differences between schools or districts seem larger than they actually were. For example, there were 11 district elementary schools that had the same mean score (41), yet they had 8 different percentile ranks that ranged from 48 to 68. In other words, apparent differences in ranks may be due to chance or have no practical consequence. Thus, the impact of this procedure was to produce school ranks that appeared meaningful but were in fact misleading.

Impact of Fallible Evaluations

The analyses by Berk and Fetler and the evaluation of the FNRTP ranking procedure make clear that there are both conceptual and practical reasons that produce fallible and misleading school ranks. The issue about the belief in the fallibility of an educational evaluation system and its relationship to job performance has been addressed by Murnane and Cohen (1986). They reviewed compensation systems that intended to alter the behavior of employees to meet the goals of the organization and concluded that when employees believe judgements about their effectiveness are made on the basis of fallible data, employees may respond in unpredictable and potentially destructive ways or the reponses may appear productive relative to other employees, but in fact may be inconsistent with organization goals. For example, employees are not likely to believe that ranking can effectively differentiate between good and superior employees. The impact of this belief was exemplified by the following response of a principal: The principal ranked a teacher as excellent, but excellent was not the highest ranking. The teacher believed she deserved the highest ranking, and when did not receive it stated, "If that's all you care, then that's all you'll get." The principal subsequently reported that the teacher's performance declined. In other words, the ranking and subsequent labeling of an employee's job performance impacts the performance of employees.

Impact of Labeling

The relationship between labeling and job performance as a function of ranking was addressed by Fetler. He indicated that schools would be labeled as a result of rankings; high ranking schools would receive positive labels and low ranking schools would receive negative labels. Attention to the lowest ranked schools can be demoralizing to school personnel, particularly when there is no direct relationship between the school's score and the efforts of the school. In their analyses of the effects of state-mandated testing programs, Anderson and Pipho (1984) remarked that although it is assumed that pointing out the successes of the education system may generate pride and serve as an example to others; it is important to keep in mind that people are motivated in different ways and what is assumed may not necessarily occur. For example, if emphasis is placed on raising a school's test score in a specific academic area, incentives to accomplish goals in less emphasized areas may be reduced (Murnane & Cohen, 1986). Thus, the ranking of schools through the FNRTP may result in many employees who do not feel valued and who may over emphasize instruction in a particular academic area at the expense of other equally important curricular areas.

Summary

In summary, conceptual and practical reasons were provided that indicated that school ranks based on the test scores of students are inappropriate. Conceptually, the piece-rate compensation used to assess factory workers' performance is not applicable to school performance because there are characteristics of students, over which teachers have little control, that influence the test scores of schools. Practically, the ranks generated by the FNRTP are fallible. The response of school personnel to being ranked on fallible data will likely result in behavior changes that are inconsistent with the goals of a school. Thus, it is recommended that the provision that requires ranking be removed from the FACET legislation.

References

Anderson, B., & Pipho, C., (1984). State-mandated testing and the fate of local control. Phi Delta Kappa, 66, 209-212.

Berk, R., (in press). Fifty reasons why student achievement gain does not mean teacher effectiveness. Journal of Personnel Evaluation in Education.

Fetler, M., (1986). Accountability in California Public Schools. Educational Evaluation and Policy Analysis, 8, 31-44.

Medley,D., Coker, H., & Soar, R., (1984). Measurement-based evaluation of teacher performance: An empirical approach. New York: Longman.

Murnane, R., & Cohen, D., (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. Harvard Educational Review, 56, 1-17.

Pencavel, J., (1977). Work effort, on-the-job screening and alternative methods of renumeration. In R. Ehrenberg (Ed.), Research in Labor Economics: Vol. 1. (pp. 225-258). Greenwich, CT: JAI Press.

Seiler, E., (1984). Piece rate vs. time rate: The effect of incentives on earnings. Review of Economics and Statistics, 66, 363-375.

RANKING SCHOOL DISTRICTS ON THE BASIS OF STATEWIDE RESULTS: IS IT MEANINGFUL OR MISLEADING?

Thomas R. Guskey
University of Kentucky

Over the past decade a growing number of states have initiated statewide testing programs as part of broad-based educational efforts. While some states administer well known, nationally-normed, standardized achievement tests, most employ tests that are developed internally, either by local testing experts or through contractual arrangements with a reputable test-development firm. The results from these tests are interpreted in a variety of different ways and used for a variety of different purposes (Pipho, 1988). In many states, however, they are used as a basis for making comparisons among school districts or schools in an effort to evaluate the quality of educational programs. Such comparisons are typically made by first rank-ordering districts or schools on the basis of some aggregate of test outcomes, and then using these ranks to determine who is doing "best" and who is doing "worst" relative to others within that state. Often such rankings are also employed in various policy initiatives directed toward improving the quality of educational programs.

While the statistical methodology for analyzing the results from such tests is sound and, in most cases, appropriately applied, the interpretations of these analyses are frequently flawed. The purpose of our study is to illustrate how the use of rank-ordering procedures based on statewide test results is far more complicated than what might appear, and that often these complications lead to inappropriate or misgiven conclusions. We illustrate how such rankings can vary greatly depending upon what outcomes from the testing program are used, how those outcomes are aggregated, whether or not input variables are used to make statistical adjustments to the outcomes, and what particular input variables are employed in making such adjustments. The implications of these differences for policy makers are also discussed.

Background

Like most states, the Commonwealth of Kentucky has initiated a variety of educational reforms in recent years. Spurred by reports such as *A Nation at Risk: The Imperative for Educational Reform*, published by the National Commission on Excellence in Education (1983), the Kentucky General Assembly enacted a series of bills, all targeted at improving the quality of education provided to children in Kentucky public schools. In fact, in 1984 alone the General Assembly passed 43 pieces of legislation bearing on education. This legislation covered a variety of areas, including academic performance, accountability to the public, professional development, community involvement, school funding, and statewide unity of educational purpose.

One of the most prominent legislative initiatives of 1984 was Senate Bill 169, which called for the identification, development, and assessment of skills deemed essential for all of the Commonwealth's students:

> The 1984 General Assembly ... enacted a bill calling for the development of essential skills that should be taught and mastered by students in reading, writing, spelling, mathematics, and library use, including research and reference skills. The bill called for annual testing of these skills in every grade. For young children who do not master those

> skills, the General Assembly provided $16 million in new funds for remedial assistance beginning in the 1986 fiscal year. (Biennial Report, 1983-85, p. 9).

The passage of this bill led to the development of a new statewide test, called the Kentucky Essential Skills Test (KEST). The KEST was developed under contract with CTB/McGraw-Hill of California, and was intended to serve a dual purpose. It was to provide a basis for determining the mastery of a defined set of essential skills, and also produce estimates of student performance referenced to national norms.

The actual construction of the KEST took place in two phases. During the first phase, in 1985, essential skills tests were developed for the domains of reading and mathematics. In the second phase, in 1986, tests were developed for the other three essential skill domains of writing, spelling, and library use. These tests were organized into 13 levels, one for kindergarten and one for each of the grades 1 through 12.

State law required that the KEST be administered to all children enrolled in public schools within the Commonwealth between the 150th and 160th days of the school year. Results from the tests were then reported to the public in the form of a statewide profile, together with individual profiles summarizing results from each of the Commonwealth's 178 school districts. These results included both criterion-referenced and nationally norm-referenced indices.

The principal normative scores used in reporting the KEST results were Normal Curve Equivalents (NCE's). The NCE's are standard statistical scores, derived by dividing the normal distribution into 99 equal segments. Scores can range from 1-99, with a mean/median of 50 and a standard deviation of 21. They were used in this instance primarily because they lend themselves to mathematical manipulations while other more common normative scores, such as percentiles, are not appropriate for such manipulations.

With the report summarizing the results from the first administration of the KEST in 1985, the State Department of Education also published a ranking of all school districts in the Commonwealth. This ranking was based upon a Total Battery NCE Score, which was obtained by statistically combining the subtest scores for reading, writing, and mathematics, and then calculating a weighted average of this combined score across grade levels 3, 5, 7, and 10.

In 1986 the Center for the Study of Testing, Evaluation, and Educational Policy (CSTEEP) at Boston College was awarded the contract to conduct an independent evaluation of the KEST Program. One of the findings from this evaluation was that educators throughout the Commonwealth saw the ranking of school districts on the basis of KEST results as one of the most troublesome aspects of the program. Most regarded the ranks as "pernicious, misleading, fostering unhealthy competition between school districts and detrimental to morale." (CSTEEP, 1986, p. iii). For these reasons, the Department of Education did not rank the school districts when the results from the 1986 KEST administration were made public. But immediately upon release of the test results, newspapers throughout the Commonwealth published articles which included similar rankings (e.g. Roser, 1986).

Some of the strongest criticism of this ranking of districts came from poorer and rural school districts in the Commonwealth which serve economically disadvantaged populations (Wagar & Estep, 1986). School personnel in these districts emphasized that it was inherently unfair to compare their test results with those from districts which have greater financial resources and serve a more advantaged student population. They argued that demographic factors such as these have a very strong influence on school achievement, but lie outside the direct control of a school district's teachers or school administrators.

It was within this context that we undertook the present study. Our initial purpose was to explore the influence of selected demographic factors on school achievement as measured by the KEST, and to determine the effect of correcting for such factors upon the ranking of school districts within the state. In particular, we wanted to determine if correcting for certain demographic factors would lead to a more meaningful interpretation of the test results and a more equitable system of evaluating the educational programs in various school districts. We then turned to consider the effects of using different aggregates of outcome variables and how this might influence the relative standing of districts as well.

Method

The data for this study were drawn from the Kentucky Essential Skills Test: Statewide Testing Results (1987), an annual report published by the Kentucky Department of Education. This report lists the results from the KEST in profiles for each of the 178 school districts in the Commonwealth, as well as data on several demographic and school-based indices. While the test data in this report were from the spring 1987 administration of the KEST, the demographic information was based on school district profiles gathered in 1986. Together, these achievement and demographic measures composed the data set for the study.

The demographic and school-based indices contained in the report included the following:

A. Percent economically deprived (%DEP) - The percentage of children eligible for free school lunch benefits in proportion to the total number of school age children in the district.

B. Percent local revenue (%LOC) - The percentage of a district's total revenue that is received from local sources.

C. Assessed property value (PROP) - This is calculated by dividing the total assessed property value in the district by the average daily attendance for the district. For convenience, this index was further divided by 1000.

D. Cost for instruction (COST) - The cost per pupil for instruction, calculated by dividing the total amount spent for instruction by the average daily attendance. The total amount spent for instruction does not include federal programs.

E. Current expenses (EXPS) - The district's annual current expenses divided by the average daily attendance for the district. It includes costs for administration, instruction, attendance services, health services, pupil transportation, operation of plant, maintenance of plant, and fixed charges.

F. Percent LEA support (%LEA) - The amount of money provided by the Local Education Agency (LEA) in addition to state funds as a percentage of the total expenditures for instructional salaries.

G. Local financial index (LFIN) - This index is calculated by dividing the local revenue per pupil in average daily attendance by the assessed property value per pupil in average daily attendance. It is intended to be an index of the amount of effort a community puts into the support of its schools based upon its ability to pay.

H. Pupil-teacher ratio (P/T) - This index is calculated by dividing the enrollment at the school building level by the number of classroom teachers reported on federal and state salary schedules.

I. Attendance rate (ATTD) - The aggregated days attendance divided by the aggregated days membership, expressed as a percentage.

J. Dropout rate (DROP) - The percentage of pupils enrolled in grades 7-12 who dropout of school during the school year. It includes pupils who became 16 and dropped out, pupils excused from school because of mental or physical disability, pupils discharged, and pupils excused from school because of marriage.

In addition to these ten demographic and school-based indices, five achievement variables were drawn from the report to use in our analysis. These included:

K. Total battery NCE score (ACH) - The average NCE score obtained by statistically combining the subtest scores in reading comprehension, writing, and mathematics for pupils in grades 3, 5, 7, and 10, weighting each grade level NCE by the number of pupils tested at that grade. This is the measure that was used as a basis for originally ranking school districts by the State Department of Education and has been used since by the popular press throughout the Commonwealth.

L. Fifth grade total battery NCE score (ACH5) - The average NCE score obtained by combining the subtest scores for reading comprehension, writing, and mathematics for all fifth grade pupils tested. This measure was included in order to explore achievement differences reflected in the elementary program only.

M. Tenth grade total battery NCE score (ACH10) - The average NCE score obtained by combining the subtest scores for reading comprehension, writing, and mathematics for all tenth grade pupils tested. This measure was included in order to consider the influence of the secondary program in the district.

N. Reading comprehension NCE score (READ) - The average NCE score obtained by combining the reading comprehension scores of pupils in grades 3, 5, 7, and 10, weighting each grade level NCE by the number of pupils tested at that grade. This measure was used to explore differences among districts based on the reading program only.

O. Mathematics NCE score (MATH) - The average score NCE obtained by combining the mathematics scores of pupils in grades 3, 5, 7, and 10, weighting each grade level NCE by the number of pupils tested at that grade. This measure was used to investigate differences among districts based on mathematics achievement alone.

Also included in the data set were the number of students in each district tested at grades 3, 5, 7, and 10, as well as the total number of students enrolled at each of these grade levels. These measures were included to determine the possible influence of district size, or if any irregularities in testing practices might have influenced test results.*

---

* Some districts were suspected of enhancing their test results by simply testing a smaller portion of the students enrolled.

Procedure

For our purposes we first replicated the ranking procedure that had been employed by the State Department of Education. This involved computing a Total Battery NCE Score for each school district in the state, and then using this score to rank-order all 178 districts. We then compared the ranks obtained by this procedure with several other alternative rankings, each employing a different aggregate of outcomes or outcomes adjusted statistically for various combinations of input variables. Specifically, we reranked the school districts based upon scores obtained from single grade levels (fifth and tenth), scores obtained from single subject areas or domains (reading comprehension and mathematics), scores adjusted for the entire set of demographic and socio-economic variables, and scores adjusted for the single most influential input variable (percent of the student population considered economically deprived). We then compared the results from these various rankings in terms of consistency and interpretive regularity.

Results

The first step in our analysis was to compute means and standard deviations for all of the measures. These are displayed in Table 1, and show there is great variation across districts in nearly all of the measures. The average percent economically deprived, for example, is 36%. But districts range from one with less than 3% to another with more than 82%. The only measures that are exceptions are the local financial index and attendance rate, where there is relatively little variation across districts.

Table 1

Means and Standard Deviations of Various School District Measures
(n=178)

Measure	Mean	Standard Deviation	Range
% Economically Deprived (%DEP)	36.01	15.96	2.7 - 82.1
% Local Revenue (%LOC)	17.81	10.07	2.7 - 66.8
Assessed Property Value (PROP)	101.10	41.12	33.9 - 268.3
Cost for Instruction (COST)	1477.03	230.00	1071.6 - 2998.4
Current Expenses (EXPS)	2040.70	309.40	1630.0 - 4522.0
% LEA Support (%LEA)	22.01	6.95	8.2 - 53.6
Local Financial Index (LFIN)	.43	.21	.14 - 1.16
Pupil-Teacher Ratio (P/T)	19.39	1.81	12.1 - 22.7
Attendance Rate (ATTN)	94.76	1.01	91.3 - 96.6
Dropout Rate (DROP)	3.34	1.90	0.0 - 13.8
Total Battery NCE Score - Combined Grades (ACH)	57.51	3.61	50.3 - 70.8
Total Battery NCE Score - 5th Grade (ACH5)	56.15	4.53	45.9 - 71.1
Total Battery NCE Score - 10th Grade (ACH10)	54.46	4.76	43.6 - 71.1
Reading Comprehension NCE Score (READ)	54.13	3.19	46.8 - 65.9
Mathematics NCE Score (MATH)	58.76	4.07	49.6 - 72.1

Next we computed correlation coefficients among the measures. These are illustrated in Tables 2, 3, and 4. It can be seen in Table 2, as might be expected, that most of the demographic variables are significantly correlated with one another. Attendance rate and dropout rate are the only exceptions, perhaps because of the rather limited variation in these two measures. However, these two variables are also probably the most alterable through the efforts of educators, regardless of the demographic characteristics of the school district or the student population served by that district.

Table 3 shows the correlations between demographic variables and the selected student achievement indices. These figures support the contention that such demographic factors are indeed related to variation in student achievement scores. The one exception is pupil-teacher ratio, which is uncorrelated to any achievement measure. Again, however, this is probably due to the rather small variation found in this index across the majority of school districts. The size of the district and the proportion of students tested were also found to be unrelated to achievement scores, indicating that while irregularities in testing procedures may exist, there is no evidence to suggest such practices are widespread.

Correlations among the selected achievement indices are shown in Table 4. These figures illustrate that although all correlations are statistically significant, many are of relatively modest magnitude. For example, the correlation between fifth grade achievement (ACH5) and tenth grade achievement (ACH10) is only .44.

Table 2

Intercorrelations Among Selected Demographic Variables
(n = 178)

Meas	%DEP	%LOC	PROP	COST	EXPS	%LEA	LFIN	P/T	ATTD	DROP
%DEP	1.00									
%LOC	-.64*	1.00								
PROP	-.63*	.77*	1.00							
COST	-.33*	.75*	.61*	1.00						
EXPS	-.28*	.72*	.61*	.95*	1.00					
%LEA	-.40*	.79*	.63*	.78*	.73*	1.00				
LFIN	-.32*	.74*	.21	.63*	.58*	.61*	1.00			
P/T	.06	-.41*	-.33*	-.73*	-.67*	-.60*	-.40*	1.00		
ATTN	-.43*	.26	.22	.13	.13	.15	.16	-.10	1.00	
DROP	.45*	-.34*	-.28*	-.07	-.07	-.22	-.20	.04	-.51*	1.00

* p < .001

Table 3

Intercorrelations Among Selected Demographic Variables
and Achievement Indices
(n = 178)

Meas	%DEP	%LOC	PROP	COST	EXPS	%LEA	LFIN	P/T	ATTN	DROP
ACH	-.65*	.52*	.46*	.36*	.31*	.30*	.28*	-.07	.49*	-.42*
ACH5	-.50*	.49*	.38*	.38*	.36*	.32*	.32*	-.18	.37*	-.39*
ACH10	-.60*	.49*	.43*	.31*	.26	.25	.29*	-.05	.37*	-.30*
READ	-.69*	.52*	.47*	.37*	.31*	.31*	.29*	-.10	.41*	-.44*
MATH	-.49*	.44*	.39*	.31*	.24	.27	.24	-.04	.48*	-.41*

* $p < .001$

Table 4

Intercorrelations Among Selected Achievement Indices*
(n = 178)

Meas	ACH	ACH5	ACH10	READ	MATH
ACH	1.00				
ACH5	.81	1.00			
ACH10	.76	.44	1.00		
READ	.92	.75	.70	1.00	
MATH	.91	.76	.64	.79	1.00

* All correlation coefficients are statistically significant at $p < .001$

Our third step was to conduct a regression analysis on the total achievement scores. This technique allowed us to eliminate or partial out that portion of the variation in the achievement scores that is attributable to demographic and school-based indices. The residuals resulting from this analysis represent achievement measures that are adjusted, or "corrected," for demographic and socio-economic differences among districts. In other words, these adjusted achievement measures are unrelated to any of the demographic factors. They are the achievement scores that would be anticipated if all districts had the same demographic and school-based characteristics.

In our initial regression analysis we used eight of the ten demographic and school-based indices included in the State Department of Education report as independent variables. Attendance rate and dropout rate were excluded because we believed these were more appropriately considered alterable outcome measures that, as mentioned earlier, are under the direct or indirect influence of educators. The results from this multiple regression analysis are illustrated in Table 5. In total, the eight variable model accounted for 45.7% of the variance in Total Battery Achievement scores. That is, nearly 46% of the variance in the total achievement scores is attributable to differences in the demographic and socio-economic characteristics of the districts.

To develop the most efficient model, we next reduced the number of independent variables included in the analysis in stepdown fashion, checking the statistical significance of the reduction in the R-square value with the deletion of each variable. A single variable, percent economically deprived (%DEP) was found to be the most powerful. In fact, the addition of other variables in the model made no significant contribution. As the results in Table 6 illustrate, nearly 42% of the variance in the total achievement scores can be explained with this single variable.

Table 5

Summary of Regression Analysis: Achievement Scores on Total Model
(Y = ACH)

Source	df	Sum of Squares	F-value	Prob	R-square
Model	8	1082.15	19.08	.0001	.457
Error	164	1162.58			
Total	172	2244.73			

Table 6

Summary of Regression Analysis: Achievement Scores on Percent of
Students Economically Deprived
(Y = ACH)

Source	df	Sum of Squares	F-value	Prob	R-square
% Econ Deprived	1	943.08	123.93	.0001	.417
Error	171	1301.64			
Total	172	2244.73			

Our final step was to rank order all school districts in the Commonwealth using various criteria, and then to compare these rankings. We first ranked all districts using the Total Battery Achievement scores (ACH), the criterion initially used by the State Department of Education. Then we ranked the districts on the basis of fifth grade achievement alone (ACH5), tenth grade achievement alone (ACH10), reading comprehension scores (READ), and mathematics scores (MATH). Next, using the residuals from the regression analysis of the full 8-variable model, we ranked the districts according to adjusted or "corrected" total achievement scores (R8ACH), fifth grade achievement scores (R8ACH5), tenth grade achievement scores (R8ACH10), reading comprehension scores (R8READ), and mathematics scores (R8MATH). Lastly, we did the same with the residuals obtained from the regression analysis using only the single independent variable, percent economically deprived, which yielded five additional rankings. Thus, we had 15 different rankings of the 178 school districts in the Commonwealth, each based on a different criterion.

To compare these different rankings we calculated Spearman rank-order correlation coefficients among them. These are shown in Table 7. If rankings were comparable across the various criteria, we would expect these correlations all to be quite high. However, this was not the case. Although the correlation coefficients are all positive, they range in value from .14 to .95, the highest being restricted to those based on residuals calculated from the same criterion variable. The median correlation coefficient among the different rankings is only .56. The dotplot in Figure 1 further illustrates the variation in these correlation coefficients.

Table 7

Intercorrelations Among Rankings Employing Different Criteria (n = 178)

Criteria	ACH	ACH5	ACH10	READ	MATH	R8 ACH	R8 ACH5	R8 ACH10	R8 READ	R8 MATH
ACH	1.00									
ACH5	.91	1.00								
ACH10	.91	.78	1.00							
READ	.80	.72	.77	1.00						
MATH	.76	.70	.64	.46	1.00					
R8ACH	.70	.57	.64	.53	.44	1.00				
R8ACH5	.55	.45	.56	.82	.19	.66	1.00			
R8ACH10	.40	.32	.33	.14	.73	.56	.15	1.00		
R8READ	.56	.65	.47	.44	.36	.86	.55	.48	1.00	
R8MATH	.58	.40	.75	.47	.30	.81	.59	.42	.60	1.00
R1ACH	.74	.61	.66	.58	.47	.92	.61	.53	.80	.76
R1ACH5	.56	.45	.56	.85	.21	.59	.91	.14	.51	.54
R1ACH10	.44	.34	.34	.17	.75	.54	.14	.95	.46	.38
R1READ	.62	.68	.50	.49	.38	.80	.53	.45	.95	.57
R1MATH	.60	.42	.77	.49	.33	.77	.56	.41	.58	.95

```
                            : .
                      .     : :           .
.                 . :.: : . : :.. .       : .       : .
:.. :.     ..::.: :::::.:.:::::::.::: :.:::.:.: :.  : :
-+---------+---------+---------+---------+---------+-----
0.15      0.30      0.45      0.60      0.75      0.90
```

Spearman Rank Order Correlation Coefficients

Figure 1. Dotplot of the Intercorrelations Among Rankings (Number of comparisons = 105)

To estimate the magnitude of the differences in these various rankings, we next calculated difference scores. That is, we simply computed the differences between each pair of ranks. The mean of these differences, of course, is equal to zero. The standard deviation of these differences, however, offers an appropriate descriptive statistic of the magnitude of rank differences.

In Table 8 are displayed the standard deviations of the differences in ranks based on various criteria. These figures show that using different criteria result in ranks that are, in most instances, dramatically different. The standard deviations of the rank differences ranged from 15.3 to 67.0, the smallest differences again resulting from comparisons between ranks based on residuals calculated from the same criterion variable. The median difference standard deviation is 47, indicating that overall, a change in ranking criterion would result in a change in rank of nearly 50 places for one-third of the school districts in the Commonwealth. Given there are only 178 school districts in the Commonwealth, a difference of this magnitude is substantial. The range of these rank differences is further illustrated in the dotplot in Figure 2.

Table 8

Standard Deviations of Differences in Rankings
Employing Different Criteria
(n = 178)

Criteria	ACH	ACH5	ACH10	READ	MATH	R8 ACH	R8 ACH5	R8 ACH10	R8 READ	R8 MATH
ACH	0.0									
ACH5	21.7	0.0								
ACH10	21.5	33.4	0.0							
READ	32.0	37.9	34.7	0.0						
MATH	35.0	38.9	42.6	53.0	0.0					
R8ACH	38.9	46.4	42.7	49.3	52.9	0.0				
R8ACH5	48.1	53.5	47.8	31.1	64.7	42.3	0.0			
R8ACH10	54.7	58.6	58.1	67.0	37.1	46.8	66.4	0.0		
R8READ	46.0	41.8	51.5	54.0	56.7	27.0	48.2	51.1	0.0	
R8MATH	46.0	54.8	35.5	52.5	59.2	30.7	46.2	54.2	44.7	0.0
R1ACH	36.2	44.4	41.1	46.7	51.8	19.6	45.0	48.7	31.3	35.0
R1ACH5	47.4	53.4	47.8	27.9	63.9	45.8	21.4	66.7	50.4	48.6
R1ACH10	53.2	57.7	57.5	65.4	35.7	48.0	66.6	15.6	52.0	55.7
R1READ	43.9	40.0	49.9	51.6	55.7	31.6	49.5	52.4	16.2	46.3
R1MATH	44.6	53.8	34.0	51.1	58.0	33.6	47.7	54.4	45.8	15.3

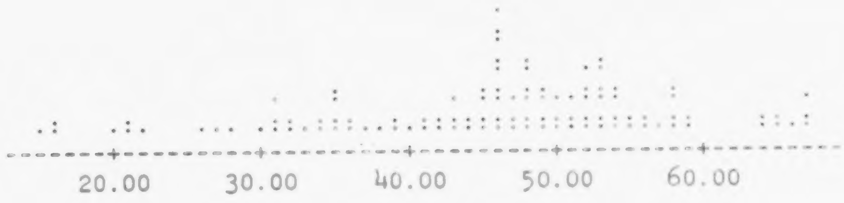


Figure 2. Dotplot of the Standard Deviations of Differences in Rankings (Number of comparisons = 105)

Cautionary Notes

Before discussing the conclusions drawn from this analysis, a few cautionary notes are necessary. First, it is important to keep in mind that the criteria used in ranking the school districts in all phases of our analysis were achievement scores derived from a test of essential or basic skills. Clearly, using such a test as the sole criterion of the quality of a school district's educational program is inappropriate. Although we would agree that these skills should be taught and that all students should master them, we would hope that the curricula of all districts in the Commonwealth go far beyond the narrow range of learning outcomes assessed by this test.

Second, although correcting for the influence of demographic factors and then ranking schools and school districts accordingly is a widely used technique for identifying those considered to be "effective," we believe such a procedure to be inappropriate. Especially in terms of the skills measured by a test of this type, we believe that students from poor families should achieve at essentially the same high level as students from more affluent families (see Brookover, 1987). We employed the technique in this study only to illustrate some of the striking discrepancies that result when districts are ranked, regardless of the criteria employed.

Third, our analysis considered only the variation that exists between school districts. Other studies have shown, however, that greater variation can exist among schools within a district than does between districts, and even greater variation still among classrooms within a school (Madaus, Kellaghan, Rakow, & King, 1979; Rakow, Airasian, & Madaus, 1978). Thus, had the unit of analysis been schools or classrooms, undoubtedly there would have been greater variation in the measures of each criterion. Nevertheless, it seems unlikely that this increased variation would result in any greater consistency in ranks determined by different criteria.

Finally, in all of our analyses we used weighted averages of test scores that were aggregates either over content areas (eg. ACH5 and ACH10), grade levels (eg. READ and MATH), or both (eg. ACH). This was done to keep our analyses parallel to those which were used to arrive at the rankings originally developed by the State Department of Education. However, we do not advocate the use of such aggregates.

Super-aggregates such as these tend to mask more than they reveal. School districts with very different patterns of achievement can have average total achievement scores that are the same, but results which suggest different paths to take if they are interested in improvement. For instance, one district could be high in mathematics but low in reading, while another district is high in reading but low in mathematics. Although their averages might be similar, the changes they might wish to make should be very different. Even a score in a specific content area such as mathematics may be, in fact, too broad to provide the kind of information needed to implement new and presumably better instruction. Patterns of results within a content area -- we are doing well in metric measurement but not as well as we would like in solving word problems -- provide a better picture of a district's performance and, therefore, better notions of the changes that should be made.

Conclusions

Our analyses show clearly that the use of aggregate test scores to rank school districts can be problematic and potentially misleading. The ranks change, in some cases drastically, as the criterion changes, and whether or not adjustments are made to the criterion. Different approaches lead to different rankings; different rankings lead to different interpretations.

In order to rank schools or school districts, one must assume that both the criteria used and the technique employed are defensible. But since both the criteria and the technique are arbitrary, any one ranking will reflect poorly and unfairly on a significant portion of the units being ranked. Therefore, it follows that interpretations based on any such rankings will be unfair, inaccurate, or invalid.

There is, however, a more fundamental problem associated with rankings such as these. In essence, they do not even serve their intended purpose, which is to tell various interested audiences how well or how poorly schools or school districts are doing. State department and school district personnel, educational leaders, thepress, and the public need good and accurate information if discussion about education are to produce positive changes in schools. Information that is potentially biased and likely to be misinterpreted willbe counterproductive to the any educational improvement efort.

Regardless of the overall level of performance of a tate's schools or school districts, a ranking procedure always produce a top and a bottom, with no way of knowing what either really means A bottom ranked school district may be doing quite well on certan criteria and a top ranked district may be doing quite poorly. The tenshortest players in the National Basketball Association, for instance, ae on the average pretty tall. Good information, something much more tha a ranking, is needed if good decisions are to be made.

Assessing the performance of schools is, as the literature makes clear, a complex task. However, that should not keep us from working hard to accomplish it. Most importantly, we should not be satisfied with weak attempts based on misleading information when dealing with issues so crucial as the educational experiences of the young.

References

Brookover, W. B. (1987). Distortion and overgeneralization are no substitutes for sound research. *Phi Delta Kappan*, *69*, 225-227.

Center for the Study of Testing, Evaluation, and Educational Policy (1986). *An evaluation of the Kentucky Essential Skills Tests in mathematics and reading*. Chestnut Hill, MA: Boston College.

Kentucky Department of Education (1985). *Biennial report 1983-85*. Frankfort, KY: Kentucky Department of Education.

Kentucky Department of Education (1987). *Kentucky Essential Skills Test: Statewide testing results*. Frankfort, KY: Kentucky Department of Education.

Madaus, G. F., Kellaghan, T., Rakow, E. A., & King, D. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review*, *49*, 207-230.

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

Pipho, C. (1988). The uses of test scores. *Phi Delta Kappan*, *70*, 278-279.

Rakow, E. A., Airasian, P. W., & Madaus, G. F. (1978). Assessing school and program effectiveness: Estimating teacher level effects. *Journal of Educational Measurement*, *15*, 15-22.

Roser, M. A. (1986, August 21). Test scores improve in KY schools. *Lexington Herald-Leader*, pp. A1 & A9-A10.

Wagar, K., & Estep, B. (1986, August 21). Wealthy districts fared best on skills test. *Lexington Herald-Leader*, pp. A1 & A8.

DISCUSSION I

Robert Ziomek
Cedar Rapids Public Schools

Austin school district's efforts to assess student performance in literature and history and the results of a fourteen district consortium in Washington State to develop a common core of district quality indicators portray attempts by school districts across the Nation to allay the fears of their constituencies that "our" district is not wallowing in mediocrity; and in order to prove it, this is how well we look compared to the "other guys". The protectionist syndrome ("I'am OK and it's the other guy who isn't.") infecting the Nation's school systems is fundamentally a self-inflicted paranoid response to an army of critics who are looking for quick fixes to problems that arise from societal neglect or political expediency. The excellent school systems know who they are, and so do their patrons! The good sytems know who they are also; their only sin may be that they have failed to keep their patrons informed. The poor systems definetely know who they are, and so does EVERYONE else! Yet every school system, in every State, is called on the carpet when the Secretary of Education displays the current "Wall Chart" and begins to make simplistic and misleading interpretations of suspicious and trivial indicators.

Now that the Council of Chief State School Officers and the National Assessment of Educational Progress have joined forces to develop tests that will be used to compare student achievement across the 50 States, as well as developing "quality indicators" (that sounds like an oxymoron to me), everyone is struggling to do the same, in spite of the admonitions of those who understand the inherent pitfalls and the self-doubts of those who are doing it! Daniel Koretz, senior social scientist of the RAND Corporation, in his article titled, "Arriving in Lake Wobegon: Are Standardized Tests Exaggerating Achievement and Distorting Instruction?" cautions that, "No set of achievement measures will alone tell us whether students are engaged and motivated, whether they enjoy school, and whether schools are giving them, not just some set of skills and information, but the enjoyment of learning and intrisic motivation to learn that they will need to be well educated and successful throughout their lifetimes." Susan Fuhrman, the director of the Center for Policy Research in Education, Rutgers University, notes that "our knowledge base is inadequate. It cannot support the development of indicators based on a substantiated model of the education system. Indicators are based partly on scientific knowledge, partly on common sense, and partly on politics. It would be nice to have scientifically agreed upon models of education, but don't hold your breath until we have them."

Terry Peterson, Chair, OERI State Accountability Study Group, in the preface to the report titled, " Creating Responsible and Responsive Accountability Systems" reflects that "The public may be displeased at times by the less than enthusiastic response educators might give to initial cries for accountability in education. Educator displeasure is often not about the principle that schools should be held accountable. Rather it stems from their legitimate concern about the accountability measures

selected. These measures may focus education too narrowly, may cause schools to pay attention to rote learning only, may encourage teaching to test items, or may be inappropriately used." So what do all these admonitions add up to? To paraphrase Professor Marshall Smith, Center for Policy Research in Education at Stanford, in developing indicators we have narrowly concentrated on INPUTS vs OUTPUTS and ignored PROCESSES!
Good luck Washington State!

In their zeal to prove to the Austin Public School community that their students did not fall into that group of students exhibiting a "deplorable lack of knowledge of history and literature", directed at the Nations public schools by Diane Ravitch and Chester Finn, the central administration of the district undertook a project to test all eleventh graders with literature and history test items used in a recent NAEP assessment. In spite of the amount of planning undertaken by the researchers given oversight responsibility for the project, internal political paranoia prevailed at the high school level("What if we don't look good, will we still publish the results?"). What the researchers discovered about what knowledge of history and literature their eleventh graders possess, after a series of hatchet blows corrupted the internal validity of their design, is either unknown or uncertain, as admitted by the writers.

An even more interesting revelation is the attempt made by the researchers to poll a group of university social scientists in hopes of arriving at a consensus of important content areas that should be addressed in social studies. They could'nt agree; or should I say that they did agree that everything was important! I often wonder what self-serving motives are satisfied by "professionals", like Ravitch and Finn, by attacking the public schools, and by default the teachers. Isn't it their colleagues who write the books that are used by the very system they are attacking? Isn't it this same group of individuals who prepare our Nation's future teachers? And isn't it the very same group of professionals who can't agree on what is important?

DISCUSSION II: THE NAME OF THE GAME

Peter Wolmut
Multnomah (OR) Education Service District

What is the true purpose of all this? Let me begin with the mixed metaphorical rhetoric I was going to use originally to close this discussion (with, of course, appropriate gestures):

> As we enter the last decade before the millenium, the sky is falling; and as the meteors strike us, while the locust plagues devour the earth, let us brace ourselves for the impending tsunami, with the hope that out of the consequential ashes, an elegant educational phoenix will rise.

Baron tells us that Florida assumes school employee performance will improve when a school's ranking is publicized. He then points out Florida's scary ritual of ranking schools based on differences of one ten-thousandth of one point. His anecdote about the teacher whose performance declined after it was labeled with a positive--but not highest--adjective, demonstrates that Pygmalion in the Classroom is not just applicable to the students.

Kentucky engages in another strange practice - done without thorough thought: the concept of aggregated weighted NCE scores across one to four grades as a basis for ranking. As Guskey points out, these aggregates mask the essential strands which need to be examined if a teacher/school/district is to improve, or know--believe it or not--that he/she/it is performing satisfactorily.

Guskey should receive the 1989 Diplomacy Award for the wording of his footnote: "Some districts were suspected of enhancing their test results by simply testing a smaller portion of the students enrolled". I suggest that the Kentucky aggregate score, should an unethical person wish to enhance a school′s results, leads that person to instigate a quiet practice of counseling low performing students to become drop outs! And, just as Pechman found with teacher cheating (1985), would it be justified on the grounds of keeping kids from being hurt? Just the other day at a meeting of research and testing directors, our ears were treated to the next step--a merging of the Kentucky and Florida methods in Texas, where weighted school (and district) averages are being ranked across the state. Can′t you see some good ol′ boy in Dalhart commenting to his wife that their youngster′s school is a fractional point better than a school in Palestine, with both of them feeling real good as a result? (Not only that, but Texas has created a new statistic. They divide schools into 3--count ′em--3 quartiles by merging the middle two! A colleague named this new statistic in impolite fashion. Suffice it to say that the name rhymes with tercile!)

But, wait. Perhaps I can′t see the forest for the trees. Gipps′ thorough description of similar ventures in the United Kingdom gives us an outside, but too similar perspective--a demand for institutional accountability, supported by simplistic methods of reporting which contain seeds of destruction to the traditional concept of an educational process.

Please be conscious of Gipps′ citation of the Goldstein/Cuttance proposition, if I may paraphrase, that what the children bring to school upon entering is the most important determinant of subsequent achievement--stated in a different context, but as old as our Coleman Report. Mistrust of the educational establishment exists in the UK as it does here. Their Government′s Department of

Education and Science Task Force says adjustment for this concept of what the child brings is not to be done lest schools become complacent. (But if there is no adjustment, then the results will be misinterpreted.) So just as in the US, the UK will also rank, resulting in--as Gipps says--school evaluation becoming straightforward, even simple, but losing its educational validity.

It is tempting to produce rhetoric again comparing the CSSO´s production of a baker´s dozen of indicators to the Christian and Lion emperors of Rome who play and did play their games because "that´s what the people want." But at least Blank and his CCSSO staff colleagues have been listening. Jim Popham (1987) made six recommendations, most of which appear to have been incorporated. For example, Popham stated that the number of assessment targets must be conceptually manageable. Blank reports only 15 content/process cells for the 1990 math assessment framework, a reduction of almost 60% from the 1986 design.

I have been concerned about the extensive disinformation surrounding the reporting of test results (1984; 1988). Perhaps the CCSSO methodology is correct: Let´s produce so much data that no matter how the states are ranked, everyone will be able to rationalize why "my state is better than yours!" Some of us with good media working relationships may be able to show how a simple ranking by score cannot give an accurate picture of accountability, of needs, of educational purpose. On the other hand, the kinds of bizarre reporting which occur in Florida and Kentucky will continue, much to the detriment of those who would like to see meager tax dollars spent for instructional items more worthwhile than endless defense of "why our state must do better." But let´s at least reveal the truth that all these indicators and rankings will allow everyone to pick the data that best suits his/her own purposes.

References

Pechman, E.M. (1985). Cheating on Standardized Tests: Why Does It Exist. In P.F. Wolmut & G. Iverson (Eds.) National Association of Test Directors 1985 Symposia. Portland, OR: National Association of Test Directors.

Popham, W.J. (1987). Suitable Assessment Instruments for Interstate Student Achievement Comparisons. In M. W. Strozeski & P. F. Wolmut (Eds.) National Association of Test Directors 1987 Symposia and Survey. Garland, TX: National Association of Test Directors.

Wolmut, P. (1984). Issues and Problems in Testing in Big City Schools: Teacher Training in Measurement. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans.

Wolmut, P. (1988). On the Matter of Testing Disinformation. Paper presented at the SRA, Inc., Invitational Conference, Scottsdale, AZ.

DISCUSSION III: COMMENTS TO PUBLIC SCHOOL RESEARCHERS, EVALUATORS, AND TESTING PRACTITIONERS

Glynn Ligon
Austin Independent School District

When I think over the information and conclusions presented during this symposium, the following truisms come to mind.

1. Ranking schools will never be done to everyone's satisfaction.
2. When rankings are done, they will always be misused.
3. Rankings will always result in a redefinition of a school's priorities to align them with the indicator used in the ranking.
4. Someone always loses in a ranking.
5. Rankings will always be made and used, because they appear to simplify an otherwise complex world.
6. Without rankings, decision makers and news reporters would have to study and understand issues much more than they do now.
7. We rank students constantly now whenever we report percentiles, and we have learned how to deal with that.
8. Rankings of schools are here to stay, and we need to learn how to deal with them.

There are two methods that I do like when a ranking is needed.

1. Comparison of the indicator of interest for the school of interest with the same for a group of schools that most closely mirror the demographics of that school. In other words, if there is a ranking of schools based upon demographic variables, then a school can be compared to the others that are just above and just below it on these key demographic variables.
2. Using a regression model to predict a school's level on the indicator of interest and comparing that to the school's actual level. In Austin, we use previous achievement scores, income, ethnicity, sex, pupil/teacher ratio, and other variables to predict achievement scores. These predictions are compared to actual scores to calculate a residual that represents how far above or below the predicted level that the actual scores fell.

In response to the key points in the individual papers...

1. In the United Kingdom model, parents are provided a narrative context description--an excellent idea. However, as a parent, how does one adjust the reported statistics based upon a narrative? How much does a parent adjust reported statistics?

2. In Florida, the downsides to using a single factor to rank schools are evident. Thus, the burden shifts to the local school districts to give alternate rankings on additional variables. When the state provides a single blind man to describe our elephant, we must provide additional blind men of our own to round out the picture.

3. In Kentucky, the point has been made that there is more information available beyond a simple ranking if we look for it. Adjusting for demographics when using a mastery/nonmastery statistic is problematic. With an average from a continuous variable, as with NRT scores, the statistical adjustment makes sense, but a dichotomous mastery score that divides a group into unequal halves is more difficult to predict. Beyond that, is it appropriate to adjust for demographics when the intent is to describe how many nonmastery students a district has? An adjustment would be appropriate only at the stage that a judgement is made as to whether or not a district performed in line with others of similar demographics.

4. The efforts by the Council of Chief State School Officers to clarify educational indicators is a worthy one. We must keep in mind that the concept of multiple indicators applies not just to reporting of multiple categories of data (i.e., dropouts, test scores, attendance rates, etc.), but also multiple indicators within categories (i.e., test scores: mastery of basic skills, averages on NRTs, SAT/ACT averages, etc.). CCSSO needs to deal with the issue of the quality of their quality indicators. Public school practitioners know the difficulty of collecting comparable data across campuses even with common definitions. Some campus-level auditing of the data CCSSO gets from the states would be very revealing--and helpful. Data look so clean and neat once they have been aggregated several times up to the state level and typeset neatly into a wall chart.

Symposium II

DOG-EARED REPORTS ARE THE BEST OF BREED: RECOGNIZING AND REWARDING EVALUATION UTILIZATION

For several years, Division H of the American Educational Research Association has sponsored a competition for the best example of the utilization of an evaluation report. The entry does not consist of the evaluation report, but rather a description of the processes used to ensure that the work is used. In this symposium, organized by James DeGracie, past winners were presented either to stand by themselves or be discussed after some more time had passed. Paul LeMahieu provides a model of evaluation, administrative, and policy development collaboration with respect to a teacher center. Elizabeth Tomblin discusses a transportable single school evaluation used with magnet schools. Lee Baldwin shows how Orange County has systematized evaluation based on what was first done with their Language Arts program. Glynn Ligon comments on the roles of timing and receptiveness and the commonalities in two Austin award winners. Stephanie Mitchell identifies factors which contributed to establishing district policy in the area of school-age child care.

SCHENLEY HIGH SCHOOL TEACHER CENTER

Paul LeMahieu
Pittsburgh Public Schools

PART I: INTRODUCTION

This paper describes a collaborative evaluation model used within the context of an extensive staff development effort, the Schenley High School Teacher Center. It begins with a brief description of the focus of the evaluation model, an innovative inservice program developed for secondary teachers throughout the Pittsburgh Public School system, with attention to the principles of design considered crucial to its success. This is followed by a discussion of the three major streams of evaluation activity conducted under a design that emphasized a strong collaboration among the researchers and program developers. Then, a discussion of a range of findings generated by the evaluation and the specific use made of the data by the key administrative decisionmakers in shaping policy and implementation processes is presented.

Finally, the conclusion describes in detail some of the factors that were critical in enhancing the role of evaluation research in the development and evolution of the SHSTC. This is offered, not as a blueprint for replication in other locales; but rather as a discussion intended to deepen the reader's understanding of the Center's experience and why it worked in Pittsburgh.

The Schenley High School Teacher Center Program

In the fall of 1981, the Pittsburgh Public Schools (PPS) began the development of the Schenley High School Teacher Center (SHSTC). A district wide needs assessment had been conducted to help the Board of Education establish its priorities as it entered the 1980's. This identified an urgent need to improve the quality of instruction in its secondary classrooms. It was felt that a locally designed staff development effort that integrally involved teachers in its design and implementation would be the most effective vehicle for this. A detailed two-year planning process that included a second needs assessment of over 800 secondary teachers, various task forces and committee meetings, etc. was employed to develop the structure for a professional development center (LeMahieu, 1984). The Center was to be incorporated into an existing district high school. Although once a prestigious high school, Schenley at that time was ranked number 12 in academic achievement among the district's 13 high schools. In order to staff the professional development center this school was "closed" on paper and reopened with faculty recruited from across the district. The Schenley High School Teacher Center opened in September, 1983, with a truly unique staff development program that coexisted within a fully operational and comprehensive urban high school. Secondary teachers from throughout the district, as well as their principals and vice-principals, were required to participate in the program sometime during the four years set aside for its implementation (1983-1987).

At the outset, the objectives for the Center derived primarily from the secondary teachers' needs assessment focused on three areas and offered teachers an opportunity:

1. to refine and expand their instructional skills;
2. to increase their sensitivity to adolescents as learners; and
3. to update their knowledge in their specific content area.

These objectives were used to shape the program's activities. Three objectives were subsequently added based on early teacher feedback. These offered teachers additional opportunities:

4. to design their own individual professional enrichment opportunities;
5. to learn more about district wide initiatives; and
6. to follow through at their home school on individual and interactive plans for continued professional growth.

For the Visiting Teachers (VTs) the professional development program entailed an eight-week "mini-sabbatical" built around activities that addressed the above six strands. Four times a year, a cohort of 45-50 teachers would participate in a Teacher Center cycle. While these teachers were at the Center, trained Replacement Teachers (RTs) took their place in the home schools. These

RTs were fully certified professional staff who had previously participated as a visiting teacher in the Teacher Center program. Among the resident Schenley staff were specially trained teachers who worked on an individual basis with one or two VTs from within their own content area. These teachers, called Clinical Resident Teachers (CRTs), assisted the VT in planning the lessons to be taught in the CRT's classroom, observed entire periods taught by the VT, and then provided immediate structured feedback to the VT. This sequence of planning, teaching, observing, and conferring is grounded in the district's model of effective teaching, the Pittsburgh Research-Based Instructional Supervisory Model (PRISM). This model was distilled from the work of Madeline Hunter (Davis, 1983) on effective strategies for organizing, reflecting on, and supervising instruction. For more information on the Schenley context, the reader is referred to Wallace, et. al. (1984); Bickel, et. al. (1987).

PART II: UNDERLYING PRINCIPLES OF THE PROGRAMS' DESIGN AND SUCCESS

The Center has achieved considerable success in reaching its initial professional development objectives. In this section we review elements in the Center's design, implementation, and context that contributed positively to the staff development effort. These comments are a description of four basic principles that undergird the Center's design and implementation activities which enhanced the effective use of the program's research design. These are presented below.

The Center's initial design was marked by a sustained commitment to collaborative planning. Almost two years before the Center opened its doors, the district initiated an elaborate planning process that directly involved several hundred teachers and administrators in design and decisionmaking activities. This planning phase had two stages: (1) the development of an overall plan for the Center (goals, structure, suggested components) by district-wide committees; (2) the development of a detailed set of program activities, role responsibilities, etc., by the newly appointed Center faculty and administration. The professionals who were the focus of the Center had key roles in shaping the Center.

A second principle involved an on-going commitment to data based/needs based planning and management. This disposition manifested itself from the very onset. The initial planning phase made use of the previously discussed needs assessment. It solicited the views of all secondary teachers. Developed by one of the planning committees and analyzed with the assistance of the district's research office, this assessment elicited teachers' views relevant to professional development needs across all high schools in the district. The results of the needs assessment were important in committee discussions shaping the overall goals and objectives of the Center. This data based planning approach continued throughout the Center's operations in many ways. This is most obvious in the maintenance of a program of ongoing formative research designed to regularly assess the status and quality of program implementation and to elicit suggestions for improving (making midcourse corrections) in specific areas of program activity. The data have been used to shape the original design of the Center and to continually improve its implementation.

A third principle has involved a commitment to innovation and flexibility. Planning and implementation proceeded with a strong sense that if the program was to be at all successful, the number of "givens" had to be reduced to a minimum, and the freedom to try new methods, structures, resource utilization approaches, etc., must be expanded. Some of the specific program characteristics discussed later on in this paper are direct manifestations of this commitment to innovation.

A fourth principle can be thought of as a hybrid leadership model that combined strong central leadership on the part of the Superintendent and the School Board with effective operational leadership on-site by the Center's principal, director, and faculty. Both the Superintendent and the School Board President demonstrated their ongoing commitment to the effort, even to the point of promoting and supporting site initiated leadership. This approach to leadership has had the effect of adding tangible support and influence to the Center's efforts, while enabling the program to grow and evolve in direct response to emerging needs, etc.

An important characteristic of the program related to the leadership and resource utilization factors just noted, has been the use of a "flat management model" that has drastically reduced the administrative layers separating program needs, organizational decision-processes, resource allocation and programmatic action. This approach is most typified by the regular breakfast meetings held by Center and district leadership. These were held frequently over the first three years and on a regular basis since then. The agendas for these meetings were developed by the Center's leadership. The membership of the meeting evolved over time, but always remained small (10-15), and was restricted to individuals directly responsible for key components of the program or for district resources and priorities, etc.

In combination, the four design principles of professional collaboration, data-based planning and management, commitment to innovation, and central/on-site leadership have made direct and important contributions to the evolution and success of the SHSTC.

PART III: THE RESEARCH DESIGN

The Center's program represented an extremely complex staff improvement effort with many features and role groups unique to the field of education. Because of the cost, complexity, and innovative nature of the Center, a research component was integrated into the program from its inception. This component involved three basic dimensions of research activity: program documentation, formative evaluation, and impact studies. These three streams of research will be described briefly. Although each will be reviewed as a separate entity, it is important to point out that they were not mutually exclusive. Rather, they existed as highly related and complementary research streams. Efforts aimed at one area of the evaluation research design often carried over into the other lines of work. For instance, information generated in the formative evaluation often identified issues that were further explored in impact studies. A brief description of each of the three strands of evaluation research will be followed by a discussion of the collaborative aspects of the overall design.

Documentation

Documentation activities were designed to build a detailed record of the planning and implementation processes associated with the Center. This record was developed so that others in the district and across the nation could benefit from the experiences of this program. This was especially important to the district because it was already considering the replication of this model at the elementary and middle school level. Documentation was primarily the responsibility of the evaluation unit of the Learning Research and Development Center of the University of Pittsburgh.

Documentation research objectives were pursued through a variety of methods. These included extensive field observations, interviews and surveys of participant groups (e.g., clinical resident teachers, visiting teachers, Center planners, etc.), and the collection of relevant documents associated with program development and implementation activities. Products from documentation research include an extensive archive of field notes and documents, numerous research reports, more general descriptive reports, professional presentations, and contributions to scholarly publications.

Formative Evaluation

Formative evaluation research activities have as their objective the generation of information for planners about the status and the quality of the implementation levels being addressed by various components of the program. The information was designed to be a useful resource for planners as they sought to improve the program. Formative evaluation activities were shared between the district's Office of Research, Testing, and Evaluation, and the Learning Research and Development Center.

Many of the activities undertaken as part of the documentation effort also generated formative evaluation information. That is, they were designed to be helpful in allowing the district's planners to make midcourse corrections. In addition to these activities, the district's research office undertook an extensive program of field observations and research activities designed to address formative evaluation issues. For instance, interviews, observations, and surveys were conducted.

The value of a formative evaluation program integrated into an implementation process is ultimately measured by the response of program leadership to such data. In the case of the Center, the record in this regard has been quite clear and will be discussed later.

Impact Studies

Impact studies were designed to develop a comprehensive understanding of the impact of the Center. The focus was on describing both short and long-term effects; intended and unintended outcomes. Impact research was primarily the responsibility of the district's Division of Research, Testing, and Evaluation (DRTE).

Given the complexity of the Schenley program, the DRTE employed a two-pronged strategy for addressing impact issues. First, its own staff directly carried out specific evaluation research studies focussed on critical aspects of impact or effects. To this end, studies were developed to assess the program's impact on individual teacher's knowledge, behaviors, and attitudes. Measures also have been taken of student performance and their attitudes about key aspects of the program (including students both in home schools and at Schenley).

Paralleling the program of impact studies undertaken directly by the DRTE, the director of that office invited interested researchers from surrounding colleges and universities to develop studies that could contribute to knowledge about the program's impact. These commissioned studies respond to specific areas of evaluation research defined by the DRTE.

Faculty members (and/or their students) whose research interests were consonant with one of the research areas were invited to contribute to the program of research. Two considerations (at least) were strictly enforced: (1) the definition and design of any evaluation research study had to be negotiated with the district's Director of Research and Evaluation; and (2) the final responsibility for any evaluation research project related to the Teacher Center had to rest with a faculty member. It was possible to secure research that covered a much broader array of outcomes and reflected a wider range of methodological approaches than would otherwise be possible. It also enabled the district to bring a considerable amount of evaluation research attention to bear on the program with a minimum amount of real expenditure.

Collaboration Within the Evaluation Research Design

The evaluation research design described above emphasized a strong collaborative theme within each of the three areas. Frequent meetings between Center leadership such as the school's principal, the program's director, and district administrators ensured that the data generated through the research was useful and relevant to the program needs. For instance, the process of developing the yearly documentation and formative research agenda began with a discussion of priority program issues as currently viewed by the Center's leadership. A draft of potential evaluation research activities based on these priorities was then created, shared with the leadership, and further shaped by subsequent discussions. In this way, Center leadership had a strong influence on the activities undertaken by the researchers.

As another example, members of the research team were included among the core planning group that met weekly for the first two years of the program to review major issues related to the program. (As the program stabilized, the core group met less frequently.) This core planning group included the principal of the high school, the director of the program, the superintendent of the school district, key Board administrators, and the evaluation researchers. Relevant findings were thus able to be periodically shared with the key decisionmakers responsible for the program's implementation; this allowed for immediate action based on the findings. Additionally, new areas in need of specific evaluation research activity could be identified through these discussions. It was this type of collaboratively designed, utilization-focused evaluation which permitted evaluation research to play a significant role in the design and implementation of the SHSTC. A summary of the evaluation research findings is provided in the next section.

Evaluation of the Center

An overview of some of the results of this program of evaluation research is provided in this section. More detailed discussion of results are available in other publications. What is offered here is intended to provide a general understanding of the effects of the Schenley High School Teacher Center, sufficient only for the discussion of evaluation use that follows.

Effects on Visiting Teachers - Some of the most important evaluative questions to be asked about the Schenley High School Teacher Center concern the impact of the staff development program on its primary focus, the visiting teachers themselves. One way to measure the impact of the program on the visiting teachers is to directly assess the perceptions of the visiting teachers themselves, as they completed their stay at the Center, a second is through direct observation of behavioral change.

In general, the visiting teachers' responses to an end-of-cycle questionnaire were favorable and suggest the positive impact of each of the program's components (see Table 1). For example, over time, better than 85 percent of the teachers either "Strongly Agree" or "Agree" with the statements that the PRISM training provided them with new instructional processes and should be offered to all teachers. Similarly, nearly two-thirds of the teachers felt that they were more knowledgeable about adolescents. They felt that the peer interaction among the teachers at the Center lessened their sense of isolation and helped them to develop a sense of collegiality with fellow teachers. Many teachers also discussed ways they would now incorporate elements of the PRISM model into their own style, or become more consciously aware of the elements already present. Teachers also commented on increasing their knowledge through the seminars in the content, adolescent development, and professional enrichment strands of the program.

The results described in the preceding paragraph were based on studies that employed self-report methodological approaches -- interviews, attitudinal surveys, journal keeping, and retrospective reports. Where the primary research questions have been concerned with teachers' perceptions, feelings, and attitudes, these self-report techniques have been appropriate and reasonably dependable methods of assessment. However, many of the summative questions concerning the impact of the Center have necessarily involved the measurements of behavioral phenomena such as the use of PRISM related instructional techniques, time on task, appropriate questioning strategies, and active student participation. In these instances, self-report data must be interpreted cautiously because of the many potential sources of bias (e.g., halo and Hawthorne effects, self-attributional bias, etc.) that are often endemic to one's assessment of one's own behavior. To address this and to assess the very serious issue of the maintenance of behavioral change in this area, a classroom observation study, one that utilized trained and independent observers, was conducted.

Randomly selected visiting teachers were solicited to volunteer for this study. Their classroom teaching behaviors were observed and systematically monitored in an extensive schedule of observations conducted both before and after their stay at the Center.

Findings consistent with the intended impact of the Center's program were obtained for a number of the teacher's instructional behaviors. In their teachings, the average amount of time spent in direct instructional interactions increased from 19 percent of the class period to 32 percent of the class period. Teachers exhibited a 20 percent increase in the use of modeling techniques and a 20 percent increase in the use of behavioral objectives to introduce an instructional task. Correspondingly, the amount of time spent in behavior management interactions was halved, decreasing from 4 percent of the period to 2 percent of the period.

Several changes were also noted in terms of student behaviors. The number of student disruptive behaviors decreased from 4 percent of the period to only 1 percent of the period while the overall rate of student "on task" behaviors increased from 61 percent of the period to 78 percent of the period. Interestingly, the most significant changes in the students' on task behaviors occurred with the focus of active learning (writing, reading, speaking) as opposed to passive learning (silent reading, listening). These findings support the contention that the Center has not only promoted a more planful and organized approach to instruction on the part of teachers, but that students manifest positive responses to the changes in teachers by increasing the number of learning related behaviors and decreasing the number of disruptive and off task behaviors.

Secondary Effects - The following discussion of associated Center effects pertain to what might be called "secondary effects" of the Center. Such impacts are "secondary" in the sense that they were not explicit goals of the Center, but nevertheless, were influenced by its operation.

One important area of "secondary effects" involves changes that have occurred at Schenley High School over the life of the Center. For example, positive changes in student performance have been particularly apparent in the population of students that attend the SHSTC -- the school's cumulative holding power (an index of the percentage of students entering a school as 9th graders who graduate four years later) increased from 89.6 percent in 1982-83 (the years immediately prior to the opening of the Teacher Center) to 93.6 percent during the 1985-86 school year.

School climate surveys of students at Schenley have also revealed substantial improvements in the school's learning environment since the opening of the Teacher Center. The most significant changes in the school's climate as reported by students have been related to their perceptions of their teachers' expectations. These include the expectations that students attend school, get to their classes on time, and spend the majority of their time focused on academic tasks. For example, in the 1984-85 school climate assessment, when students were asked "What do your teachers expect you to be?" 66 percent of the students indicated either "an above average student" or "a superior student." Similarly, when asked to rate the item, "My teachers are very serious about doing a good job," 86 percent of the students selected either the response options of "agree" or "strongly agree." Many, if not most of the students, seem to respond positively to the climate of high teacher expectations. Twenty-eight percent of the students indicated that they spend between 30 to 60 minutes on their homework each night; 40 percent indicated that they spend one hour or more on homework each night. It should be underscored that this level of "academic press" is a sharp increase over that reported by students prior to the opening of the Teacher Center.

Administrators' Program - One final area of Center activity involved work with secondary school principals and vice-principals. Although the Center always focused explicitly on teacher growth and renewal, the early research on maintenance of effects made it increasingly clear that building administrators must play a critical role in the effort if gains made at the Center were to be sustained and extended as visiting teachers returned to their home schools. Thus, by the beginning of the second year of the program, the Center's leadership and a core planning group of administrators developed a program designed to simultaneously strengthen secondary administrators capacities for (and interest in) sustaining teachers' growth and to stimulate and support professional growth among the administrators themselves. The general features of these efforts have been discussed in an earlier section of this paper. The outcomes from these efforts were measured in a variety of ways.

Most of the research focused on Center programs for administrators was formative in nature. That is, the research efforts were an integral component of the implementation process, with questions emphasizing how the program could be improved and what activities would be most useful in subsequent rounds of the effort. However, administrators were surveyed at the close of each experience, and these surveys did contain several questions related to what the respondent perceived to be the accomplishments of their program.

Principals expressed great concern about leaving their buildings for the amount of time required by the development effort. Indeed, both the teacher and the administrator efforts involved a considerable amount of time. Both groups expressed similar concerns about the impact their absence would have on their classrooms and schools, respectively. The issue of lengthy absence from one's workplace is a very important concern. The support structures must be carefully designed, and the benefits must clearly outweigh the costs. In the case of the administrator program the data suggest that both of these conditions were largely met.

Use of Evaluation Research Results

The research and evaluation efforts attending the Schenley High School Teacher Center influenced both the program itself as well as subsequent program development in the district. In this section, some of the ways that the evaluation information was used will be listed. Essentially, the types of use fall into three areas, changes in the program, carryover effects in the district, and "instrumental" uses. Each will be discussed briefly below.

Program Modifications - One major area of use of evaluation research results was modification of the Teacher Center program itself. There were numerous examples of this, a few of the more influential are included here. First, there was a complete overhaul of the scheduling of the program's activities (early in the program's life). For the first two cycles, the clinical teaching and review component was distributed throughout the cycle. Feedback on the difficulties associated with this model from both the visiting teachers and Schenley staff (especially related to scheduling) resulted in the third cycle's implementation of "teaching" versus "non-teaching" weeks. This model was subsequently modified for the second year of the program; the visiting teachers were divided into two groups based on their content area. These two groups were scheduled with opposing "teaching weeks" (i.e., weeks 3, 5, and 7, or weeks 4, 6, and 8).

Also, early cycles had their training in the district's teaching model (PRISM) distributed throughout the eight-week cycle. Then, the training was presented in a two-week time frame, at the beginning of each cycle. Visiting teachers reported that they were better at seeing the total picture given this cohesive presentation of each part. They were also more confident at using all the parts of the model in their subsequent teaching activities at Schenley. Clinical resident teachers also reported that the massed presentation of the instructional model enhanced their collaborative work with the visiting teachers.

As a third example of the way the schedule was modified based on evaluation results, early program participants recognized the opportunity available to them to work on their own special projects using the variety of resources made available at the Center. Some teachers in these early cycles, however, complained that their disjointed schedules often did not allow them to take full advantage of this. Blocks of time were made available to teachers to allow them to schedule this optional activity.

The development and offering of specific activities and seminars in response to ongoing assessment of the teachers' needs and desires (e.g., sessions on classroom management, discipline, and regular meetings between the visiting teachers and the superintendent) serves as a second example. Specifically, initial seminars in the adolescent development strand of the program were general discussions of broad issues, such as "Self-Concept, Self-Esteem, and Self-Discipline." Based on feedback from teachers in end-of-cycle questionnaires and discussions, this strand of the program evolved into one that presented much more specific issues. This strand of the program was also shaped by comments from the Center's Board of Visitors (a review panel of eminent educators and scholars) who identified a need for seminars focusing on variations in adolescent experience across differing racial and ethnic populations. Seminars related specifically to minority youth, yet generalizable to the broader adolescent population, were added (e.g., "Racial and Sociocultural Differences in Cognitive Learning Style").

Seminars in content area strands of the program were also changed, based on participant feedback, to become ones designed to help the teachers become more aware of current trends in their field, as opposed to learning "how to teach the subject." Feedback from the Center's Board of Visitors on the multi-ethnicity issue mentioned above was also used to modify this component to include more seminars dealing with minority youth issues. For example, in the English content area "The Role of Afro-American Writing in Contemporary American Literature" was included.

Finally, a major revision of what is called the "replacement process" was conducted based on evaluation results collected from various audiences. In this, the scheduling of cycles was rearranged so as to permit a period of collaborative planning and transition between the visiting teachers and their replacements. In response to an inspection of the replacement process, a formal strategy for collaborative planning and effective transition was developed, evaluated, and implemented.

Other Programs and Activities Initiated Within the District - In addition to modifications to the Teacher Center program itself, evaluation research was also helpful in identifying areas of need that the district then pursued through the initiation of new programs. One of the observations made within the evaluation research was that the burden of educational reform could not fall upon the shoulders of teachers alone, and, therefore, could not be effected through Teacher Centers only. For example, the research on the clinical resident teachers at Schenley led to a redefinition of instructional (or department) chairs throughout the city. This included the definition of goals (as being more closely aligned to the evaluation of instruction and the quality of content delivery) and the establishment of methods to pursue those goals, as well as the determination of the training and support necessary to see the new chairperson's role effectively implemented. This effort was strongly influenced by the research on the somewhat similar clinical resident's role at the Teacher Center.

A second example in this area is the district's adoption of an Employee Assistance Program (EAP). Research examining the VT-CRT interactions uncovered the fact that in some instances, the CRTs were faced with trying to address teacher instructional needs that were functionally rooted in personal crises or problems being faced by the VT. A pilot EAP was proposed to address this need and has since been expanded to include teachers across the district.

A final area in which the Schenley research influenced other programming throughout the city has to do with the development of Teacher Centers at other grade levels. As of this point in time, the district has implemented an elementary Teacher Center and is in the final stages of planning a middle school center, due to open in September, 1988. Both of these efforts have benefited from the knowledge gained and documented through the research on the Schenley effort.

"Instrumental" Uses of the Evaluation Research - The final area in which the evaluation results were "used" by the district was more instrumental in nature. That is, it proved useful at a number of junctures in securing support, both financial and organizational for the Schenley Teacher Center and the subsequent similar efforts at other grade levels. For example, the Board of Directors' decision to support the expansion of the Teacher Centers to include other grade levels was predicated (in no small part) upon the understanding of its effects that were derived from the program evaluation. The same can be said of both the national recognition and the extensive support of foundations for the Schenley effort. To date, the Schenley Teacher Center has been the beneficiary of approximately 1.8 million dollars generally in philanthropic support. The Brookline Elementary Teacher Center has received .5 million dollars in similar support, which was encouraged by the results documented through the Schenley research efforts.

PART IV: CONCLUSION: REFLECTIONS ON UTILIZATION OF FINDINGS

The Schenley High School Teacher Center represents a model of district-based staff development which has far-reaching implications for the design, implementation and evaluation of similar efforts. In reflecting on those factors that contributed uniquely to the use and usefulness of the evaluation at the SHSTC, five factors appear to be most important:

1. The Evaluators Were "In From The Outset" - The very first evaluation research undertaken in connection with SHSTC prefigured even the planning and development of the Center's program. It was an assessment of teachers'

their inservice and staff development needs. The results of this needs assessment were used to inform the deliberations and planning of the more than two hundred teachers responsible for the design of the Teacher Center program. From that point forward, the research activity (more particularly, the information that it produced) maintained an accepted place in the ongoing planning for the center.

2. The Evaluation Research Was Integrated Into The Planning and Management Processes - The evaluation team was represented on the program's steering committee. This was done both so that they could fulfill documentation research requirements, and so that whenever appropriate and possible, the deliberations of the planning groups could be informed by relevant research findings. Moreover, the research team made periodic reports of findings as they became available from each of the various studies. Early in the program's implementation, a week was set aside each summer for a team of teachers and others to work on modifications of the program for the subsequent year's delivery. In each instance, these program planning sessions were initiated with a review of the research findings from all of the previous year's activities.

3. The Collaboration Between People From Inside and Outside the District - The unique formulation of the research activities so as to include researchers from both within and outside of the district afforded certain benefits that contributed to the evaluation use. It contributed to the general validity of the undertakings by permitting appropriate assignment of research tasks to researchers. In addition, it permitted the establishment of sympathetic and beneficial relationships between the researchers and various informant sources and between the researchers and the program leadership. This enhanced the quality of data and findings, and in the long run contributed to its usefulness.

4. The Establishment of an Ethos of Respect for Evaluation and Its Use - Much has been written about personalities and the mesh of personalities between evaluators and their clients. While the importance of this factor is easily noted, the prescriptive direction to be taken from it is not quite so. However, when the evaluation project extends over so long a term as this one and it is formulated in the way that this one was, it is possible to take some guidance from the writings about evaluator-client relationships to enhance utilization. To be specific, the evaluators worked from the outset to build an ethos of evaluation use. This was done by refusing to recognize the typical boundaries of an evaluation project. The effort was not completed with the writing, submission, or even presentation of a report of findings. Rather, no report of results was complete until it had been made intelligible to, and useful for, its relevant audience. Being involved with the project from the outset and for the long term made this perspective on use possible. Moreover, the forms of trust and relationships that develop between evaluators and the program leadership contributed to the use. For the most part, credit here goes to the program's leadership, whose intellectual honesty imbued them with a willingness to ask hard questions and entertain and respond to difficult answers. However, it should be observed that this disposition bred a culture in which many others became willing to do the same, and evaluation use maximized.

5. The Conception of the Evaluation Activities as a "Program of Many Studies" - On reflection, it seems that this fundamental conception of the approach to evaluation contributed to the usefulness and the use of the findings in many ways. First, with a number of studies ongoing at any time (some just finishing, some just starting) there was often a set of results to review, discuss, and respond to. This is a far cry from the traditional lengthy wait

for a single set of results. In addition, this approach enabled the researchers to synchronize analyses and the availability of relevant information to the natural program and planning cycles of the project.

This "modular" approach to the evaluation activities enabled the research to respect the information needs of multiple audiences and devote greater attention to each. It also permitted the evaluation program to be methodologically sympathetic to the many research questions and the various audiences that it sought to serve.

Finally, approaching the evaluation project as a program of many studies permitted, even encouraged, greater flexibility in responding to the developing research and information needs of the program. To state the obvious, it is impossible that any single study can conceive of all of the relevant questions that will ultimately be of interest. The notion is wholly quixotic in the face of the many changes that the program underwent over time. By framing the evaluation project as a series of studies, it proved no great burden to modify particular studies, add additional analyses, even terminate certain studies or initiate new ones in response to the evolving interests or information needs of the program's leadership or other constituencies.

Beginning in 1986, district and Center leadership began to plan for what would follow the first round (Phase I) of Center activity. The essential broad goal here is to maintain and expand the momentum already begun so that on-going staff renewal and further improvements in student outcomes will continue to occur in each high school in the District. Based in large part on the evaluation findings from Phase I of the SHSTC, educational renewal, school-based staff development, and shared decisionmaking within each building have emerged as critical targets of the Phase II initiative. The challenge, of course, in all of this is to institutionalize into the other high schools in the District some of the critical features of the Center's own experiences with professional development and educational improvement.

While no one is yet quite sure where the planning for Phase II will lead, it is safe to assume that the role of evaluation and research activities will continue to be paramount in the design, planning and implementation of the next phase of secondary staff renewal in Pittsburgh. The collaborative model of program evaluation, administration and policy development provides the District with a degree of assurance that what ultimately emerges from the planning of Phase II will serve the schools in a timely, appropriate and instructionally effective manner.

BIBLIOGRAPHY

Bickel, W., Denton, S., Johnston, J., LeMahieu, P., Saltrick, D., and Young, J. (1987). Clinical teachers at the Schenley Teacher Center: Teacher professionalism and educational reform. Journal of Staff Development, 8(2), 9-14.

Davis, L. (1983, April). Improving the performance of teachers. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

LeMahieu, P. (April 1984). "An assessment of the needs of secondary teachers relevant to professional development: A story of evaluation use." Presented as a round table session at the annual meeting of the American Educational Research Association, New Orleans, LA.

Wallace, R., Young, J., Johnston, J., LeMahieu, P., and Bickel, W. (March, 1984). Secondary education renewal in Pittsburgh. Educational Leadership, 41(6), 73-77.

Zigmond, N., Kohnke, R., and Miller, S. (1986, April). Assessing the impact of teacher inservice on secondary special education classrooms. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

EIGHT INDIVIDUAL MAGNET SCHOOLS

Elizabeth Tomblin
San Diego City Schools

Introduction

Magnet schools in the San Diego City Schools have been a successful part of the voluntary integration program and have been cited by the Superior Court for their success. Until recently, the magnet programs have been evaluated on the basis of the drawing power, that is, the ability to attract and hold students, particularly students who improve the racial balance of the school. While this is certainly an important and basic criterion, the quality of the instructional program is of great importance. In a district as large as the San Diego City Schools (over 112,000 students), the more popular magnet themes are replicated at more than one school to handle the enrollment and reduce waiting lists and to create better racial/ethnic balance.

Using that basic criterion of enrollment draw as an indicator of success, one school may draw students while another of the same theme may have difficulty drawing students. Data did not exist to evaluate the various programs on any criteria other than the numbers which the school has attracted and district-wide test scores. It then appears that some magnet schools are more successful than others. What then accounts for the difference? How do we know if a school is successful? Sometimes the success of the magnet school appears to be more related to the magnet emphasis; other times the success appears to be more related to the school environment, instructional program, etc. Consequently, the Community Relations and Integration Services Division requested an evaluation of magnet schools.

Developing the Design

Discussion between the Evaluation Department and the Community Relations and Integration Services Division determined that needs dictated evaluations of individual magnet schools separately, not an evaluation of the magnet program in its entirety. Certain questions would be common to each school, such as questions about the basic skill attainment, school climate, transportation provisions, etc. Other questions would be unique to a school. For example, questions about the dance studio at a school for the creative and performing arts would be irrelevant at a computer/science magnet. The amount of commonality appeared to be greater than the differences, so the assistant director developed a "generic design" which would serve as a basis for each school's evaluation, but which could be modified for each school according to its unique program.

While this model was developed for the purpose of conducting a comprehensive evaluation of an individual magnet school program, it is adaptable for the evaluation of any integrated school program, or any specific school site or school-based program. The model has been accepted by the Court in two integration cases in two states. The design has as its purpose:

> "To determine the breadth and depth of the magnet programs in suggested areas at a specific school with respect to academic achievement in specialized magnet and regular classes and curricular offerings; to determine the attitudes of parents, students and teachers about the magnet program; and to examine the costs of the magnet program."

The design has the flexibility to be modified based upon the unique needs of a specific site. The design includes multiple sources of data and multi-methods of data collection, utilizes implementation studies and combines quantitative and qualitative methods of analyses. Timelines make it possible to be on site at the school less than a month, permitting minimal disruption to the normal school schedule, yet providing for extensive input by staff, for classroom observations, and for survey administration. Scheduling gives attention to the collection of data available centrally before the evaluation team enters the school, again minimizing any disruption to the school day while enabling the collection of quality and accurate data.

Instruments which have been developed have been field tested and validated prior to use in the schools. Statistical analyses utilize downloading to IBM/PC for manipulation using SPSS. Report preparation uses word processing for efficiency.

Cost analyses are provided; in many cases a cost-effectiveness ratio can be established. Another unique feature is the examination of grading practices by ethnicity and gender as well as by departments or types of classes.

The components of the design, instruments, methods of data collection and analyses and specific examples are available from the Evaluation Department. However the purpose of the NATD seminar at AERA was not so much to discuss the design as it was to discuss the utilization of the evaluation. The effective utilization is a result of the process as much as perhaps more than the design itself. Magnet programs evaluated using this model include a math/computer/science magnet, an elementary math/science magnet, a fine arts magnet, a school for creative and performing arts, a writing academy and a preparatory high school among others.

Process

The evaluation of an individual school is an unusual event, and a process is needed in order to involve the school staff and to inform all those who need to be informed. Evaluation can be very threatening to school staff, particularly those who are unfamiliar with the process. Teachers are often particularly anxious about the prospect of observers in the classroom, confusing the presence of a program observer with personnel evaluation.

The process creates an advisory committee of the staff members to work with the evaluation staff to tailor the design and to determine timing and procedures of data collection. The advisory group also has the opportunity to review report drafts. Following the evaluation, a procedure for solutions and responses to recommendations has been implemented. Involving the staff from the beginning gives them a chance to learn what the evaluation is to be, to have input into the wording of questionnaire items without the pressure of having to develop the instrument nor to collect the data. Their input is very helpful to assuring the smoothest data collection and the use of appropriate words which will be meaningful to students and parents. (For example, our questionnaire might use the words "Math Resource Room" when the school uses the words "Math Lab." Using the word familiar to the school insures more accurate responses.) When staff members have a chance to understand and participate in what data will be collected, they are more likely to feel some ownership in the evaluation and more receptivity to the formulation of action plans based on the data collected in the evaluation.

Along with the involvement of staff members from the beginning of the evaluation, there are people who need to be informed, such as the Assistant Superintendent who supervises the operation of the school, the Assistant Superintendent who is in charge of the magnet school program, the magnet program coordinator, and at certain times, the staff in the Educational Services Division who have the responsibilities for curricular programs.

To insure that all know the process and that the communication occurs, a Sequence of Events was developed and distributed to all concerned. The process is included as Attachment 1 to this paper.

An important part of the process is the timing of the reports to the Board of Education. The Superintendent's Cabinet receives the report when it is completed and at the same time that the school receives the final copy of the report. The report to the Board of Education is delayed. Following receipt of the report, the school has several weeks to months to develop a response to recommendations and an action plan which they will use to improve their school. The call for papers comes from the Assistant Superintendent for Community Relations and Integration Services, and the magnet coordinator works with the school to build the plan.

The report to the Board of Education is timed so that the Board receives the evaluation report and on the same day, immediately following, receives the school's plan of action. If the evaluation report were reported several months prior to the school's response, the chances are higher that the media might take the most negative parts of the report and criticize the school and that the school's response might sound more defensive than constructive. The purpose of the evaluation process, to use the cliché so many evaluators have intoned, is not to prove, but to IMPROVE.

The process was designed to facilitate improvement above all. The Board of Education takes action to approve the school's action plan or to send it back to the school for revisions. In most cases, they have approved the action plan reports as presented, but in some cases they have requested additional information or revisions to the plan.

Schools Evaluated

At the time of the utilization recognition, seven magnet schools had been evaluated using this design and this process. Two more evaluations have occurred for a total of nine during a three year period, and one is scheduled for 1989-90. Three of the schools are elementary schools, one with math magnet emphasis, one with a fine arts emphasis and one with an individualized learning emphasis. Three of the schools are high schools, one featuring a business magnet program, one with an International Baccalaureate and writing academy, and one a preparatory high school. Two other schools have multiple grades, one a school for creative and performing arts has students in grades 4-12, and a computer/math/science magnet has students in grades 7-12.

The schools' reaction to the evaluation process have varied. Most have been quite cooperative, even enthusiastic, helpful and have seen the information as valuable to improving their school program. Two schools have reacted negatively, one due to some internal problems existing prior to the evaluation and noted in its findings and one which did not understand or want to understand the evaluation process and which remained very uncomfortable with the idea that they would be evaluated. Consequently, they were defensive about and resistant to the recommendations.

To decode the initials on Attachment 1, a few words about the organization of the district is necessary. The Superintendent's Cabinet is composed of the heads of divisions. Community Relations and Integration Services Division (CRISD) operates the magnet programs (among other integration and community programs). School Operations Division (SOD) sees to the daily operation of the schools. Planning, Research and Evaluation (PRE) Division conducts planning and enrollment forecasting activities, research and evaluation studies. The Evaluation Department is in the Planning, Research and Evaluation Division.

Process for School Action Plan

Attachment 2 shows the format in which a school responds to the recommendations and provides its action plan. The process has generally been a group review by selected or volunteer staff members, followed by several meetings to discuss and formulate the school's plan. In some cases, the Evaluation Department has met with the school to review findings or discuss recommendations. In some cases, the school has proceeded on its own.

Results and Uses of Evaluation

Besides the response format presented above and the action plans by the schools, there have been a number of requests for the design and for the reports. School districts within California have called or visited the San Diego City Schools to discuss, to request a copy and to ask for permission to use the design. Outside the state, a notable example is the use of this magnet design in the presentation to the judge in the Kansas City, Missouri desegregation case as part of the plan for desegregation/integration of the Kansas City schools.

With the current interest in schools of choice, requests for information have come from people writing reports for Secretary of Education at the federal level, from the State Department in California, and from across the nation. Magnet schools seem to be the designated public school model for schools of choice.

Conclusion

As school reforms continue to be important and restructuring of schools becomes a focus nationwide, the implications for developing a design and institutionalizing a process to examine an individual school's program are far reaching. As demographics change and more districts move beyond desegregation to integration, they must assure that resegregation doesn't occur once students enter the schoolhouse door. For school districts turning to voluntary integration programs, magnet schools are proving viable and successful ways to provide quality academic and specialized programs in an integrated school setting within public schools. This design, process and concept evaluates the results and provides for effective utilization of evaluation to benefit students and to improve programs.

Attachment 1

Magnet School Evaluations Sequence of Events

Event or Task	Responsibility
Designate schools to be evaluated and order for evaluation	CRISD
Discuss selection with School Operations Division	CRISD
Notify School Operations Division, PRE and school principal	CRISD
Meet with school principal	Evaluation Department
Principal appoints advisory committee of teachers and parents to advise evaluators	School principal
Modify generic design (if necessary) for each individual school	Evaluation Department
Comment on specific design	CRISD, and School Operations Division
Prepare and distribute final design	PRE
Meet with school committee to review surveys and design	Evaluation Department
Develop/prepare instrumentation	Evaluation Department
Review final instrumentation	CRISD and School Operations Division
Collect data	Evaluation Department and school
Analyze data	Evaluation Department
Prepare report draft	Evaluation Department
Share preliminary findings with school advisory committee, staff (survey results, etc.) and district magnet coordinator	Evaluation Department
Prepare final draft	Evaluation Department
Send report through editorial review board	PRE
Send final draft to CRISD and School Operations Division and school principal	PRE

ELEMENTARY LANGUAGE ARTS

Lee Baldwin
Orange County (FL) Public Schools

The purpose of this paper is two-fold. First the program evaluation for which the award was given will be described. This will include a description of the program, a summary of the major findings and recommendations, and actions that were taken as a result of the evaluation. The second purpose of this paper is to describe the system of program evaluation which has been developed to institutionalize the utilization of program evaluations.

Part 1: Elementary Language Arts Program Evaluation Results

The program evaluated was elementary language arts, which served 68 elementary schools and 2 exceptional education centers. The purpose the language arts program is to provide experiences to students for them to develop writing, reading, listening, and speaking skills. The program emphasizes the process approach to writing, a method which is gaining in support. The method stresses the process of composition as well as the need to think about what is to be written, to confer with classmates and the teacher, to rewrite, and to enjoy the process in the finished product. The program was directed by a district elementary program consultant for language arts with one resource teacher assisting.

Use of the evaluation was evident by the action plan which was developed by the program staff; the action plan addressed every recommendation of the evaluation report. This action plan became a part of the District Plan, which provides direction for the district for a five-year period. Based on the evaluation results the program consultant established an objective to provide training for 224 teachers on how to teach "Writing as a Process." The objective became a part of the department plan for elementary education. The action plan was reviewed and approved by the Associate Superintendent for Elementary Education and the Deputy Superintendent for Instruction.

Perhaps the most significant impact of the evaluation was its influence on the decision to hire five additional resource teachers to assist in training classroom teachers in the process approach to writing. Two findings from the evaluation contributed to that decision: 1) A large majority of classroom teachers trained in the process approach to writing were effectively implementing the approach and 2) Student writing skills increased significantly more in trained teachers' classrooms than in untrained teachers' classrooms. ($p < .035$).

The evaluation report was also used to convey information from school-based personnel to district-level administrators on questions pursued in the evaluation which address the following: The appropriateness of program objectives, satisfaction with an availability of operational guidelines for the program, implementation of the program according to its design, perceptions of staff development activities, opinions on use of the single basal language series, and the availability of materials and supplies. This information was conveyed primarily through a discussion of the evaluation report with the program consultant and the associate superintendent for the program before the report was submitted to the Deputy Superintendent for Instruction and the Superintendent. After the Superintendent reviewed the report, he called a meeting with the Program Consultant for Elementary Language Arts, directors of elementary education, associate and deputy superintendents for the program, plus the evaluators and the Associate Superintendent for Planning and Governmental Relations. The Superintendent led a discussion of the findings, conclusions, and recommendations in the report and directed that an action plan be prepared to address the recommendations. He expressed the desire for further information on the program. A follow-up evaluation was conducted during the 1987-88 school year to focus on student achievement in writing as a result of resources made available to hire five additional resource teachers to assist in training teachers and implementing the process approach to writing. Another focus of the follow-up evaluation was to determine if teachers were indeed implementing the process approach to writing as it had been designed.

There were several factors involved with the evaluation which may be somewhat unique in the world of program evaluation but are institutionalized in the Orange County Public Schools. It is the system of review itself which has contributed to the utilization of this and other evaluations. The authors of the original award submission did not think that this evaluation was unique in its degree of utilization but was, rather, a good example of utilization. At the most recent convention of the American Association of School Administrators the Superintendent of the Orange County Public Schools was selected to present our program evaluation model to a meeting of school superintendents. The presentation was a part of symposium held to demonstrate methods for improving decision making.

Part 2: Program Evaluation Model in Orange County

Program evaluation exists in the Orange County Public Schools for the purpose of providing useful and timely information to decision makers so that they can make informed judgments. A decision maker is defined as any person who needs to make decisions about a program including program contacts, principals, associate superintendents, and the superintendent. What makes the program evaluation efforts in Orange County somewhat unique is the systematic process of utilizing the evaluation findings. These efforts to utilize program evaluation findings were recognized in 1987 when the Program Evaluation Office won the Evaluation Utilization Award presented by the American Educational Research Association for "an outstanding contribution in facilitating the use of evaluative information by actively and expertly organizing, conducting, and reporting such information so as to enhance its potential for use by decision makers." The award is given by the School Evaluation and Program Development Division of AERA. The award was accompanied by a $500 contribution to a scholarship fund in the Orange County Public Schools.

Program evaluation is implemented in accordance with the Comprehensive Plan for Program Evaluation. The plan outlines the sequence of events and responsibilities of personnel involved in the program evaluations and is designed to ensure that program evaluations fulfill their purpose of providing useful and timely information. The first step in the evaluation process is the development and revision of the five-year calendar. The Director of Testing and Program Evaluation solicits input from all department heads and other decision makers and recommends programs for evaluation each year to the Superintendent. Although the calendar is a five-year calendar, it is reviewed and revised annually. After approval of the calendar, evaluators are assigned to the programs they will be evaluating, and then they develop the program evaluation questions and plans in collaboration with the program contact. Evaluation questions are also approved by the Superintendent. After the evaluation questions are approved, the evaluation is conducted during the school year. A final report is written and then reviewed with the program contact.

The review process occurs in the summer and fall and culminates in the Superintendent's review of program evaluations. The program evaluation executive summaries (summaries of major findings and recommendations) are reviewed by all administrators with responsibilities for the program from the program contact up to the Superintendent. The Superintendent then holds a meeting to review and discuss each program evaluation report. At the meeting, the Superintendent charges the program contact with developing an action plan to address all of the recommendations in the evaluation. The action plans are then incorporated into the district plan. The review process and action plans ensure that program evaluations are utilized and evaluation recommendations are addressed to improve programs in the district.

What follows are excerpts from the Comprehensive Plan for Program Evaluation which describes the steps and timelines for the conduct of program evaluations (Attachment I). In addition, an excerpt is reprinted from the roles which various personnel play in conducting or facilitating the program evaluations (Attachment II). Finally an organizational chart is presented which shows the relationship of program evaluation to the organization as a whole (Attachment III).

The plan, flow chart, and timeline of responsibilities give an idea of the system and the structure for institutionalizing a program evaluation process. It is designed to create and maintain a management system in which appropriate programs are identified, meaningful questions are asked, appropriate data are collected, accurate and fair reports are produced, and recommendations are acted upon. The plans and procedures ensure that a system exists to facilitate the utilization of program evaluation.

A second key component in the management system is the actors. The roles of the Superintendent and Associate Superintendent are supportive and facilitative. The message from both is clearly one of support throughout the organization. Both commit appropriate amounts of time and other resources when needed. These resources include reviewing and approving the evaluation calendar, evaluation questions, and the evaluation reports along with providing an adequate budget. Their actions in approving the various steps of the evaluations have primarily been to give an air of importance and to ensure adequate input to the process rather than control the process.

In summary, the program evaluation system has been implemented to provide useful information to decision-makers. The organizational structure and roles within the structure are designed to facilitate the conduct and utilization of evaluation. The comprehensive plan establishes a process by which programs and issues are identified in a timely manner. Finally, the personnel who fill the roles understand their roles and act as a team to mutually implement the program evaluation effort. The interaction of an organization's structure, a comprehensive plan, and committed role players together leads to an effective management system for program evaluation.

Attachment I

Sequence Of Annual Events To Conduct A Program Evaluation

The following sequence of events is provided to guide all personnel involved in evaluations as to responsibilities and time frameworks.

A. Selection of Programs for Evaluation: The Five-Year Calendar (*Early February*)

Due to the large number of programs in the district and the constraints on staff time, the first critical event for evaluations is a wise choice of programs to be evaluated. The Program Evaluation Office will prepare annually a recommended five-year calendar of programs to be evaluated. The programs for the upcoming year will be reviewed in light of events that may affect evaluation priorities. The decision on which programs to evaluate requires a broad knowledge of the school system. Judgment will be based on input from the Superintendent and his administrative staff, the School Board, programs emphasized in the District Plan, state and federal mandates, student testing results, other external and internal evaluations, information on program emphasis at the school level from the review of decision packages which are submitted as part of the planning process (EPRMS), and input from community organizations. Some programs may need evaluation assistance that spans several years.

The proposed calendar will be recommended by the Associate Superintendent for Planning and Governmental Relations (hereafter referred to as AS/PGR) and adopted by the Superintendent. Programs selected in February for an evaluation will be evaluated during the following school year. The five-year calendar should allow sufficient time for appropriate staff to submit program objectives.

B. Development of Questions to be Pursued by the Evaluation

It is critical that evaluators become familiar with programs and know the important issues and decisions that need to be made about the programs. An evaluator, i.e., a member of the Program Evaluation Office, will be assigned to each program to be evaluated. To ensure that appropriate questions are pursued, evaluators will do the following:

1. Receive initial input from the Superintendent and the administrative staff. This will be accomplished in conjunction with receiving input for developing the five-year calendar. This input will allow evaluators to know issues and concerns that need to be addressed and to share them with program staff at the outset. This input will help ensure that evaluations focus on the most important issues and address questions about which decision makers are most concerned. (*End of February*)

2. Meet with experienced and new program consultants, who are designated as program contacts for the evaluations, to introduce the program evaluation process, share experiences from having participated in an evaluation, and begin formulating plans for the upcoming evaluations. (*April*)

3. Become thoroughly familiar with the program and other evaluations conducted in the area through literature searches; available documents, such as curriculum guides and handbooks; and discussion with the program contact and other appropriate personnel. The evaluator needs to become steeped in the existing program and in research and evaluations related to the program to know how to effectively plan the evaluation. (February through July)

4. Receive a list of program objectives and activities central to the accomplishment of these objectives from the program contacts. This requirement of a list of objectives and activities will alert program staff to the necessity of having objectives crystallized in measurable terms and will provide initial information to evaluators to familiarize them with the program so that they can begin to formulate evaluation designs. (*End of April*)

5. Develop the evaluation questions to be pursued for each program in consultation with the program contact; include information received from deputy and associate superintendents. (*June*)

6. Present the evaluation questions to be pursued to the Superintendent, deputy superintendents, AS/PGR, and other appropriate associate superintendents for their approval. The evaluation questions will provide communication to ensure that important issues have been included and interpreted as intended and that the district's leaders approve of the approach to evaluations being pursued. (*July*)

C. Design of the Evaluation (*April through July*)

The evaluator will design evaluations in consultation with program staff and with consideration to input received as described in B above. The Planning/Report Form for Program Evaluation, Parts I-V, will be utilized as a guide. Parts I-V of the form include program description, objectives, evaluation questions, assessment instruments, and procedures. As the evaluation is planned, decisions will be made on the type and format of reports for the individual schools.

A Program Evaluation Monitoring Checklist, which includes timelines and responsibilities, and a Cost Analysis for Program Evaluation will be completed jointly by the evaluator and program contact and approved by the appropriate supervisors. These documents will ensure that responsibilities for tasks and costs are mutually understood. A cost analysis of an evaluation is important to make personnel aware of what is involved to carry out the plans. Otherwise, many expenses may be overlooked. The design of an evaluation, for example, sampling instead of collecting information from an entire population, may be adjusted after projected costs are realized. Expenses incurred for evaluation purposes will be the responsibility of the Program Evaluation Office. Expenses for the program per se are to be paid from the work location budget of the program.

D. Selection of Evaluation Instruments *(April through July)*

Evaluators will decide upon instruments to be used and then design instruments as needed. The Testing Office will advise on achievement data currently available and its applicability to program evaluation, as well as on test selection, interpretation, and development. Test data currently available will be utilized to the fullest extent.

E. Presentation of Plans to Advisory Committees *(May)*

Advisory committees will be convened to give their input, when necessary; for example, to discuss the compatibility of timelines with school schedules and the feasibility of evaluation plans.

F. Distribution of Plans and Timelines *(Mid-August)*

Evaluation plans and timelines of activities will be presented to all principals prior to preschool planning. These plans may be presented through a memorandum or at regularly scheduled meetings for principals.

Evaluation plans and timelines of activities will be presented to program staff at the preschool staff development sessions. This presentation of evaluation plans will allow program staff to know the kinds of information about their programs that will be gathered. If no staff development sessions are planned for a group, another meeting time will be scheduled.

G. Collection and Analysis of Data *(Throughout school year following A-F above, September-May)*

Data collection and analysis will be the responsibility of the Program Evaluation Office. The most effective approach will be decided upon individually for each program.

H. Preparation of Evaluation Reports *(June of the year following A-F above)*

Evaluation reports will be prepared by the evaluator. The content of the reports will follow a format which includes a table of contents, an executive summary, purpose of evaluation, related district goal(s), program description, program objectives, evaluation questions, assessment instruments, procedures, findings, conclusions, and recommendations. Before the reports are prepared for release, they will be discussed and shared with the appropriate associate and deputy superintendents and program staff, who will have the opportunity to react and give suggestions. Reports will then be reviewed by the Director of Testing and Program Evaluation and the AS/PGR and submitted to the Superintendent.

I. Distribution of Evaluation Reports *(August of the year following A-F above)*

The complete evaluation report, including an executive summary, will be bound and copies distributed to the associate superintendent with general supervision over the program, the program contacts and their supervisors, and other designated individuals as appropriate. Executive summaries only will be sent to School Board members, the Superintendent, deputy superintendents, associate superintendents who do not receive the complete reports, the

Assistant to the Superintendent for Community Relations, and principals for the applicable level. The total report will be made available upon request. For an evaluation there may be other designated individuals who will receive selected information. The choice of personnel to receive information will be based on involvement with the program and usefulness of the information. Findings will be presented to school-level personnel at a staff development session or other appropriate meeting early in the school year.

J. Development of and Follow-up on Action Plans *(August and September of the year following A-F above until implementation of action plan)*

Because a program evaluation is expected to be a useful tool for making decisions about the program, an action plan will be developed to address the findings and recommendations of the evaluation report. The recommendations will define what is needed; how the need is to be addressed will be left to the appropriate program staff as they develop the action plan. Budget decisions and policy implications will be a part of action plans. A continuing objective of the District Plan will be to develop and implement action plans for programs which were evaluated in the previous program evaluation cycle. The development of action plans will include the following:

1. Evaluation reports are prepared by the Program Evaluation Office and approved by the Superintendent.

2. Superintendent meets with program contacts, their supervisors, directors, associate superintendents, and evaluation personnel to review conclusions and recommendations.

3. Superintendent charges program personnel to prepare action plans, including timelines, to be approved up through deputy superintendents. A copy of the actions plans will be sent to the AS/PGR. The AS/PGR will consider, in conjunction with deputy and associate superintendents, components of action plans to be included in the District Plan. Associate superintendents will consider components of action plans for departmental and individual objectives; therefore, the action plans need to be developed before departmental and individual objectives are finalized.

4. Action plans will serve as monitoring instruments to be utilized semiannually in conjunction with gathering information for the District Plan's progress report. Action plans include the following: recommendations, response/action to be taken to include activities, personnel responsible, timeline, and progress to date.

Attachment II

Roles of Various District Staff In Program Evaluations

In the foregoing sequence of events for an evaluation, the roles of various personnel were cited. These roles are summarized as follows:

A. Superintendent

1. Provides input for establishment of five-year calendar; reviews and adopts calendar
2. Provides input on major issues and concerns to be addressed by evaluations
3. Approves final list of evaluation questions to be pursued
4. Reviews and reacts to evaluation reports and approves executive summaries
5. Reviews recommendations and conclusions of evaluation reports and charges program personnel to prepare action plans
6. Transmits executive summaries to School Board members and community groups

B. Superintendent's Administrative Staff (appropriate associate and deputy superintendents)

1. Give suggestions for which programs should be included on the evaluation calendar each year including school-based programs
2. Apprise evaluators of the issues and concerns that need to be addressed for each program
3. Review and approve the evaluation questions to be pursued for each program
4. Support the evaluations conducted in their areas by working closely with directors and program contacts involved in evaluations in their areas
5. Review and approve action plans based on findings of evaluations

C. Associate Superintendent for Planning and Governmental Relations

1. Supervises the coordination of system-wide program evaluations and communicates between the Superintendent's administrative staff and the Program Evaluation Office on policy decisions about program evaluations
2. Integrates program evaluation with overall planning process to recommend in conjunction with deputy and associate superintendents the components of action plans to be included in the District Plan

3. Recommends proposed five-year calendar for program evaluation to Superintendent
4. Carries out VII-B, 1-3

D. Director of Testing and Program Evaluation

1. Prepares annually a recommended five-year calendar for program evaluation
2. Assigns program evaluation responsibilities to staff
3. Recommends evaluation questions to be pursued to the AS/PGR
4. Designs and conducts program evaluations
5. Approves evaluation designs
6. Supervises implementation of program evaluations
7. Informs appropriate personnel of evaluation plans and timelines
8. Reviews evaluation reports; recommends reports to the AS/PGR
9. Directs distribution of evaluation reports
10. Coordinates the distribution of school-level evaluation reports
11. Assists in integrating program evaluation with overall planning process

E. Evaluator

1. Recommends programs to be continued beyond one year for the five-year evaluation calendar
2. Becomes familiar with programs being evaluated, important issues to be addressed, and decisions to be made about the programs
3. Develops evaluation questions to be pursued as outlined in VI-B
4. Designs evaluations in consultation with program staff and advisory committees, where appropriate; the design includes the selection and/or development of assessment instruments
5. Implements evaluations, including collection and analysis of data
6. Presents results to program staff
7. Prepares final evaluation reports
8. Prepares school-level reports of evaluation results, where appropriate
9. Reviews progress in implementation of action plans

F. Directors and Supervisors of Program Contacts

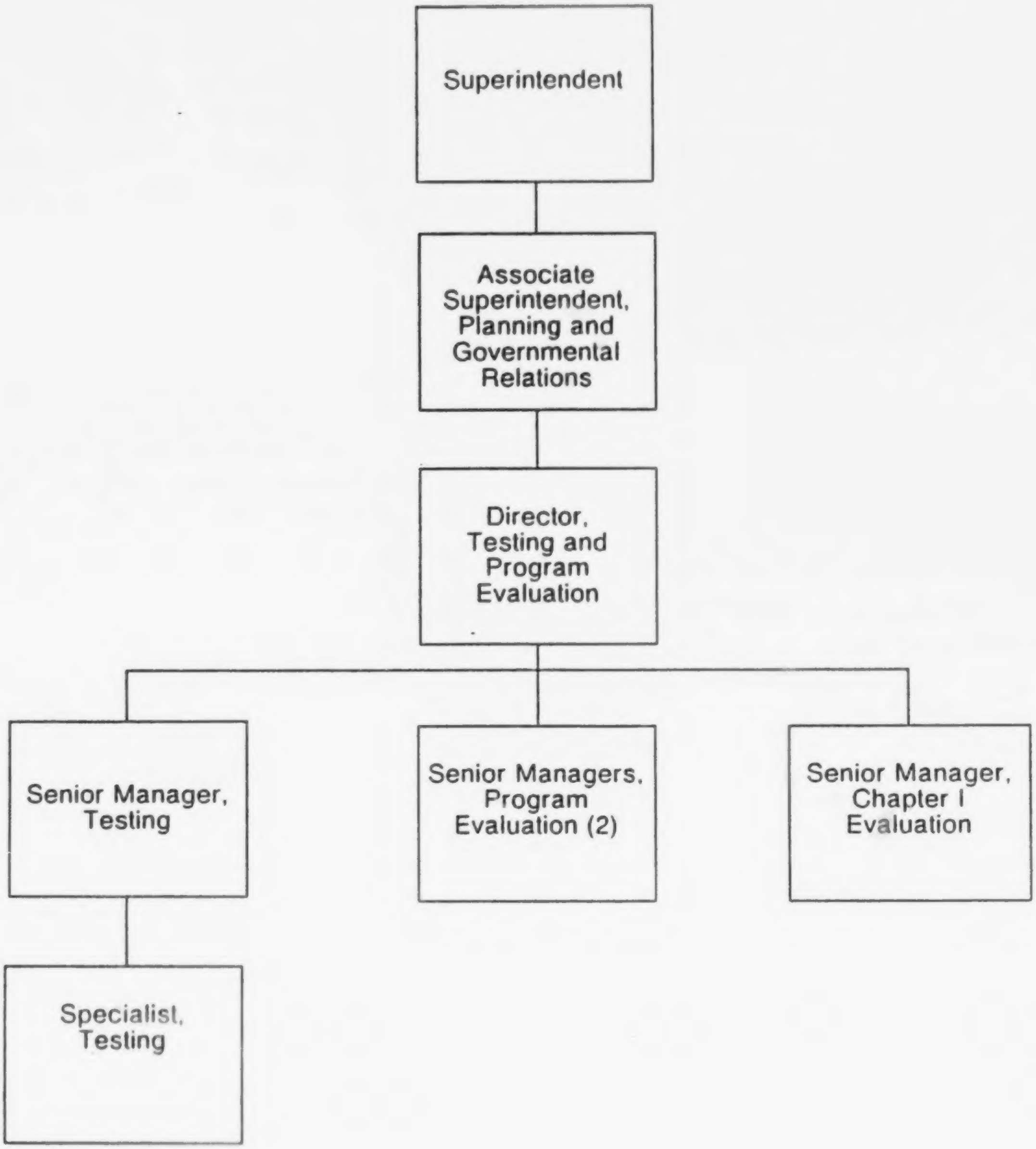
1. Apprise evaluators of issues and concerns to be addressed
2. Review timelines and responsibilities for tasks and costs of the evaluations
3. Review evaluation reports before they are prepared for release
4. Review action plans

G. Program Contacts

1. Provide in writing the objectives of the program in measurable terms and essential activities to achieve the objectives
2. Familiarize the evaluator with the program
3. Work with the evaluator to design the evaluation, select instruments, and plan the timeline and costs
4. Assist in implementation of the evaluation as agreed upon with the evaluator
5. Read and give reactions to a draft of the evaluation report
6. Develop an action plan based on findings
7. Implement the action plan

Attachment III

Organizational Chart Testing and Program Evaluation


Superintendent
Associate Superintendent, Planning and Governmental Relations
Director, Testing and Program Evaluation
Senior Manager, Testing
Senior Managers, Program Evaluation (2)
Senior Manager, Chapter I Evaluation
Specialist, Testing

TEAMS MASTERY
and
DISTRICT DROPOUT PREVENTION EFFECTIVENESS

Glynn Ligon
Austin Independent School District

My opinion, at least as of today, is that evaluation utilization is 45% timing, 5% context or the ability of the organization to respond, 5% perseverance by the evaluator, and 45% willingness of the program staff to respond. This places much of the success of utilized evaluations external to the reporting process. I do not believe that an eager evaluator can force an evaluation to be useful. The reports recognized by Division H of AERA for utilization have benefitted from a receptive context. However, I also believe that without the insights and efforts of a skilled evaluator, even the most potentially useful of evaluation reports cannot fulfill that potential.

The two reports in this symposium from the Austin (TX) Public Schools are prime examples of the necessity of a receptive and appropriate context. Our dropout report emerged at the beginning of a local, state, and national focus on dropouts and a demand by an anxious public for improvement. I could and would argue that our report influenced that cry for action, certainly locally. Our TEAMS mastery report also benefitted from a local, state, and national demand for basic competencies on the part of high school graduates and an expanding emphasis on statewide testing programs.

Both of these reports created or contributed to an atmosphere of conflicting priorities for our schools. Dropout prevention efforts caught schools between the demand for lower dropout rates and the demand for better discipline and higher attendance rates. The emphasis on improving our District's TEAMS mastery rates caught schools between this priority and the longer term pressure to score high on our standardized tests. In a broader sense, the teachers found themselves torn between emphasizing the broader curriculum associated with the norm-referenced test and the more focused curriculum measured by the TEAMS, a criterion-referenced test.

Therefore, both reports required our district to discuss and clarify goals, and both TEAMS mastery and dropout prevention received higher priorities than their competing goals.

TEAMS mastery and dropout prevention both created a team spirit, a cooperative atmosphere between evaluators and program staff. Uncharacteristically, the evaluators were seen as part of the solution rather than just as the messenger of bad news. This relationship led to a major caution that must be heeded.

Bias by the evaluators must be monitored as ownership in programmatic activities recommended in the evaluation reports develops. In other words, the evaluators must be very cautious that they can still evaluate the programs objectively once they have exercised such influence and have created a stake for themselves in the success of the program activities. Fortunately, as individuals, our evaluators have resisted the ownership trap and still view themselves as observers rather than active partners in the educational process. However, in other school systems, the philosophy of evaluations lends itself more readily to ownership in programs designed around recommendations of the evaluators.

The two individual papers detail the actions taken by the evaluators to ensure that the 5% of utilization that is dependent upon the evaluator was successful. Each evaluation truly represents an example of evaluators responding to the information needs of the educators well beyond the covers of the published, formal report.

TEAMS Mastery

Evangelina Mangino, David A. Doss, and Glynn Ligon

The Program

The Austin Public Schools has implemented a major instructional effort to ensure the mastery of Texas' Essential Elements as measured by the statewide Texas Educational Assessment of Minimum Skills (TEAMS) test. The program is for students from grade 1 through graduation who do not have the skills in Texas' Essential Elements to perform at a mastery level on the TEAMS in mathematics, language arts, and written composition. The program provides matching of local curricula to TEAMS objectives, campus-level goal setting, practice/diagnostic tests, tutorials, two-minute lessons targeted at single TEAMS skills, etc. Resources for the program come from State Compensatory Education funds, Federally funded sources such as Chapter 1, and local funds. The program, although multifaceted in approach across the grade levels, is clearly focused upon one goal--to have a higher percentage of students in Austin master the TEAMS test than in any other urban district in Texas. (Note: For Austin, this is an appropriate goal, because we have the lowest percentage of low-income students among the eight urban districts.)

Use of Evaluation Findings

To convince people that there is a problem. First and foremost, the community, Board of Trustees, the administration, and the teaching faculty had to be convinced that Austin students really do perform less well on the TEAMS tests than students in other urban districts. This was a bigger challenge than would be apparent at first glance because the assumption had always been that Austin students would easily outperform all others. One of the first to be convinced was the Superintendent, who quoted the results in almost every speech, presentation, and written communication he made to District as well as community groups. Enclosed is a poster that he had made and sent to every school and department--"AISD's Challenge--Be #1 in Urban TEAMS Scores."

The Office of Research and Evaluation published a formal annual report on student achievement with particular emphasis on an analysis and evaluation of the relative standing of Austin students on the TEAMS. In addition, an oral presentation with graphics was made to the Board of Trustees in a televised public meeting. Then individual school reports were sent to each principal giving each schools' mastery percentages compared to the previous year and all other District schools. Upon request, the testing staff spoke to groups of principals and school faculties with more detailed interpretation of the results.

To motivate people to make improvements. Second, the District had to be so convinced that Austin students should be performing higher that the number one goal of the District would become to have the highest mastery percentage among the eight urban Texas districts. The Board of Trustees adopted as the District's number one priority in 1986-87 "To improve student performance--especially TEAMS mathematics." Attached is a list of the priorities with this one being number I.A. The divisions of elementary and secondary education directed their principals to set school goals and that one of them had to be the improvement of mastery percentages on the TEAMS.

To focus resources on specific problem areas. Third, the instructional staff had to learn from the achievement report and associated analyses where the students needed to improve. To supplement the diagnostic reports provided to all districts by the Texas Education Agency, the Office of Research and Evaluation designed and produced classroom, school, and District summaries by individual objective. Overall, the clearest need was in mathematics from grades 3 through 7, but other areas were also highlighted--especially for individual campuses. From these analyses, the instructional staff, centrally and on each campus, was able to focus their attention and resources on the areas they determined to be the highest priority--either in terms of the lowest skill levels or the areas of greatest potential for improvement.

To monitor progress toward improving student achievement. Finally, at the request of the Superintendent, custom designed reports were produced to compare schools' ranks within the District from year to year, and to show the schools' differences between their mastery levels and those of the District and the State. With much local concern focused on sacrificing higher order skills by the concentration on the basic TEAMS skills, the Office of Research and Evaluation identified those schools which had achieved outstanding TEAMS gains as well as gains on the Iowa Tests of Basic Skills (ITBS) and Tests of Achievement Proficiency (TAP). By examining these comparisons, instructional staff identified schools with consistently good performance and began to study what made them successful and what could be shared with other schools.

Unique factors in this evaluation contributing to its utilization.

LOOP. Linking Outcomes to Organization Planning (LOOP) is a management tracking system developed by AISD's Office of Research and Evaluation within the Department of Management Information, specifically to identify priorities that require attention and careful monitoring. The Superintendent asked for a system such as LOOP to ensure that critical evaluation findings become a part of the District's planning look rather than remaining isolated findings that lacked impact. Attached is a description of LOOP and how it ensures that key finds are addressed and monitored. Also attached are copies of the memos from the Superintendent to staff assigned the TEAMS priority.

To make LOOP work, the Superintendent and Cabinet review key evaluation findings and other information about the District at an annual planning retreat. A set of priorities is drafted and passed through a review process that culminates in adoption of the official priorities for the District by the Board of Trustees. The priority for improving TEAMS performance was assigned to the two assistant superintendents for elementary and secondary education. They and their staff developed their strategies for improvement, outlined them for the Superintendent through LOOP, and reported on their progress periodically.

LOOP provided the planning, monitoring and assessment structure that ensured that the evaluation findings were used.

In-Depth Analyses. This evaluation was distinguished from those of other districts by the incredible depth to which the analyses stretched. The usual dimensions were explored--grade, ethnicity, income, etc.--but any other idea that an instructional person offered was also investigated by the evaluation/testing staff--mobility during the school year, program participation, years in the District, instructional grouping, feeder patterns, etc.

If someone asked and it was possible, the Office of Research and Evaluation conducted the analyses. With such a high priority on improved test scores, the testing staff catered to almost every whim of the instructional staff. Two representative examples of the analyses follow:

The District redrew school boundaries for the 87-88 school year and invalidated the standard TEAMS reports provided by the TEA. The reconfigured neighborhood schools are quite different from the paired schools with cross-town bussing. The testing staff used the computerized student assignment files to reanalyze all summary reports and student listings for the

new school assignments. This tremendous effort allowed schools to set goals based on their new student populations. As part of the statewide accreditation monitoring program, all students were ranked on TEAMS, with the lowest 25% and 5% targeted for special monitoring. In Austin, the original rankings became inappropriate, but the reanalysis allowed the District to convince the TEA to use the current student population for rankings and to discard the original ones.

Instructional staff and the Superintendent requested that schools with a history of outstanding TEAMS performance be identified as models for others. Again computer analyses were compared across the three years of TEAMS to highlight grades within schools that had been consistently high. However, to balance the fact that TEAMS measures a restricted, low range of skills, the schools' performance on the ITBS was also analyzed to find the campuses that had successfully achieved high mastery on TEAMS without sacrificing overall gains on the ITBS.

Cooperation and Coordination Between Evaluation and Instructional Staff. Seldom has an evaluation staff and instructional staff worked closer together to ferret out every bit of useful information from a set of data. In fact, the evaluation staff at times provided analyses with the disclaimer that instructional staff had overstepped the bounds of reasonable dissection of data for diagnosis. Several frank discussions occurred when the appropriateness of the use of the data was vigorously debated. The end result has been not only a thorough analysis targeted at the questions that the instructional staff had, but also a better understanding on the part of the instructional staff of the limitations of data--and on the part of the evaluation staff of the instructional staff's thirst for diagnostic information.

* * * * * * *

District Dropout Prevention Effectiveness

David A. Doss and Freda M. Holley

The Program

For several years our district had been reporting dropout information to the state. However, we had reason to believe that the information was not accurate. Furthermore, the results did not take into account students who dropped out over the summer. Therefore, we undertook a study to describe the dropout situation for our district. Specifically, we evaluated both the ability of our educational system to hold students until they graduate and the validity of the dropout statistics being reported.

Use of Evaluation Findings

From the beginning, the most important use of the findings has been to alert the district and the community to the fact that Austin has a dropout problem. It was a shock to many (especially in 1983) that about one quarter of our students drop out within four and one-half years of becoming 14. The results have been included in numerous newspaper and television news reports since that time.

The publication of the report led to a series of dropout initiatives by the district beginning in the months following its spring, 1983, release and continuing today. The major impacts are described below. Some resulted directly from the study. Others came later as the outgrowth of previous activities. For example, the Evening High School resulted from a recommendation of the dropout prevention task force.

1982-83

a. Report issued.

b. Superintendent asked for the development of a plan for annual reporting of dropout rates by campus.

1983-84

a. Dropout prevention was adopted as one of the major district goals for the 1983-84 school year by the Superintendent's Cabinet.

b. A dropout prevention task force made up of representatives from campuses, central office, and the community was created. The task force created a dropout prevention plan (Attachment A), which was presented to the Board of Trustees in February, 1984. Information from the student was included in the report. Recommendations included the creation of an evening high school and a school to serve as an alternative to suspension.

c. The annual dropout rate reporting procedure was put into place and the first results reported. Attachment B is an example of the kind of document sent to the schools.

d. ORE provided Houston ISD with dropout study results to aid them in planning their court-mandated study of dropouts.

1984-85

a. In February, 1985, the Board of Trustees adopted dropout prevention as one of three priority areas to be addressed by the system's accreditation plan for the period 1985-86 through 1989-90. A copy of the plan is enclosed as Attachment C. Evaluation findings were used in preparing the plan.

b. The Evening High School began operation.

c. An alternative school to serve in lieu of suspension began in a rented home.

d. For the first time, each school adopted a dropout prevention plan.

e. A part-time dropout prevention coordinator was appointed.

f. A Dropout Prevention Coordinating Council was appointed to assist in coordinating resources for high-risk students.

g. The annual reporting of dropout rates continued.

1985-86

a. In the years following their publication, the results of the study were shared with other evaluators in a number of ways. Papers were given at AERA, results were shared with other evaluators in Texas through the Joint Urban Evaluation Council, and Spectrum, the journal of the Educational Research Service, published a modified version of the report. Because of AISD's experience with dropout studies, David Doss was appointed to serve as a member of the Advisory Committee for the statewide study of dropouts funded by the Texas Department of Community Affairs, and Freda Holley served as a member of the Task Force on Collecting National Statistics on Dropouts of the Council of Chief State School Officers. Recommendations from each group have endorsed a dropout definition which is essentially the equivalent of Austin's.

b. Schools added the development of a dropout prevention goal to the required dropout prevention plan.

c. An elementary school was vacated to serve as the location for the enlarged suspension center.

d. The Austin Chamber of Commerce appointed a Dropout Committee to bring the support and help of the business community to bear on the problem. The student results were presented to the committee.

e. Annual rates were reported again.

1986-87

a. The Board of Trustees adopted a dropout prevention policy for AISD (Attachment D). The evaluation of district efforts is required by the policy.

b. AISD's dropout research was featured in Deborah Burnett Strother's PAR article on dropouts in the December, 1986 issue of the Phi Delta Kappan. David Doss and Glynn Ligon served as members of the source panel for the article.

c. The dropout prevention coordinator was appointed full time to the position.

Unique Factors in This Evaluation Contributing to its Utilization.

The single most important factor contributing to the utilization of the findings was timing. The report was rushed into print so that the results would be public information in time to be included in a series of newspaper articles on dropouts. The report also came just ahead of the nationwide reemergence of interest in dropouts. Because the staff and the community had not seen many recent articles or new reports on dropouts and dropout pevention, the information and the challenge it presented were fresh. As interest rose within the community, the availability of the results of the original study and the annually updated rate (with the attention surrounding its reporting) probably assisted community members in keeping the emphasis on dropout prevention before the administration and School Board.

Being a little ahead of the national trend also meant that the methods we used in counting dropouts were ready for consideration by other districts and agencies as described above.

A second factor in the utilization of the results was the style in which the report was written. It departed dramatically from previous (and subsequent) Office of Research and Evaluation (ORE) reports in the degree to which an editorial position was taken and the extent to which an attempt was made to draw an emotional response from the reader.

- The title, Mother Got Tired of Taking Care of my Baby, was personal, memorable, quotable, and in some quarters volatile. It was distinct from the cold, technical titles of most reports and was meant to attract readers and draw them into the report.

- The report uses student art on the cover and has illustrations thoughout (Attachment E). The cover was selected to suggest turmoil.
- "Case studies," which were composite descriptions based on interview responses, were included to personalize the plight of the dropouts (see Attachment E).

We attempted to make the dropouts real people to the readers and to convince them that the problem deserved their attention.

A third factor is a fortuitous change of assignment for Freda Holley, the instigator, prodder, and shaper of the study. During the summer following the report's release, she left her position as Director of Research and Evaluation to become the Assistant Superintendent for Secondary Education. Her change in position placed her where she could facilitate the task force formed to create solutions. When task force recommendations emerged, she could foster the recommended solutions such as the evening high school and the alternative education center. It is unlikely that the dropout issue would have disappeared completely, but the changes in district program would probably have come more slowly and would have followed the national interest in dropouts rather than preceding it.

The fourth factor important in the utilization was capacity--the capacity of the Office of Research and Evaluation to do the study and the capacity of the district to respond. ORE had the resources (the people, the money, and the data files) to do the study and the respect of the rest of the district so the results were taken seriously. Also, the study came to the district "out of the blue." It provided unanticipated, disconcerting information, and it came with no strings attached. That is, it was not done in response to pressure from an external group or agency. Arising as it did as a concern from within the district, it may have met with less resistance than other findings.

At the time, the District was enjoying the success of a smooth transition to a new desegregation plan and looking for new challenges. The financial situation was comparatively strong. Equally significant results might not lead to the same level of activity today.

The fifth factor was the sharing of the results with professional colleagues. As described above, the results were shared in both informal and formal ways. The findings have literally been disseminated around the world. The impact that the findings have had on other evaluators or researchers is unknown, but potentially, it may have been widespread.

The final factor that influenced the impact of the study was its comprehensiveness. Not only were dropout rates calculated by sex, ethnicity, etc., but attempts were made to predict dropping out from centrally available information, and 100 dropouts were found and interviewed. The study provided a data base that has been valuable over the years for providing answers for other questions as they arise. For example, we have looked at the impact of desegregation and course selection at the ninth grade on dropping out. Had we not had the data to fall back on, interest in the dropouts might have flagged somewhat.

SCHOOL-AGE CHILD CARE

Stephanie Mitchell
Portland (OR) Public Schools

Over the past decade, increased emphasis has been placed on the use of evaluation in educational decision making. Different techniques have been used to encourage evaluation utilization in school districts with varying degrees of success. The basic premise underlying the efforts to improve evaluation use is that use of data will enable school leaders to become better decision makers. This should result in improved educational policy decisions.

Efforts to improve the way schools acquire and utilize evaluation information are based on two assumptions: that the evaluative data are adequate and relevant and that they are available to the appropriate decision makers. This paper explores these two assumptions and looks at how one district's policy makers acquired and utilized this body of evaluation information. Similarities and differences in evaluation acquisition and utilization for school and district-based decision makers are explored and implications are developed for how further evaluation utilization could proceed based on this analysis.

The focus of the paper is the impact of evaluation utilization on the development of a district policy regarding school age child care. Findings from the *1985-86 Evaluation Report on School Age Child Care in the Portland Public Schools* (Moilanen and Mitchell, 1986) and the historical context of the evaluation utilization process are presented. The paper first investigates how evaluative data were used to help develop a policy statement and guidelines for school age child care, and then examines the process of building commitment to evaluation utilization among District policy makers. Lessons learned in this process are discussed in an effort to provide guidance to other researchers interested in evaluation use and its impact on policy development.

Prior Research

There has been a dramatic expansion of investigations into evaluation utilization in recent years (Alkin and Stecher, 1983; Glasman, 1986 and 1988; Pechman and King, 1986). Investigations have focused on evaluation-based leadership, the use of evaluation in educational planning and decision making, and studies of the acquisition and use of evaluative data. Alkin (1983), Braskamp and Brown (1980), and Glasman (1986, 1988) provide a general synthesis of this research as it relates to the utilization of evaluation research by teachers, principals, school districts, and educational programs. This research has also been synthesized around adapting evaluation use to differences among users (Newman, 1980).

Research on evaluation utilization has also resulted in training manuals and guides that focus on characteristics of the evaluator, the user, and the organization or program (Alkin, 1985). Specifically, the user's interest in evaluation, commitment to use of data, openness to change, and initiative in seeking out data for decision making have been found to increase the chances for utilization. In terms of characteristics of evaluators, their commitment to use, rapport with users, credibility, willingness to involve users, and political awareness seem to be key factors in facilitating successful evaluation use. Organizational factors related to uses of evaluation may include the role of the program and its interrelationships with other programs in the district, interactions between evaluators and users, the relevance and specificity of evaluation recommendations, and the frequency, timing, and manner of presentation of the evaluation information reported to the program (Pechman & King, 1986; Ripley, 1985).

The School Age Child Care Evaluation Study

In 1985 the Board of Education and Superintendent in Portland, Oregon requested an evaluation of the District's school age child care programs to help them make better informed decisions about the possible continuation or expansion of before and after school child care options. Staff of the District's Research and Evaluation Department were asked to study the need for before and after school child care, examine the school's role and responsibility, and analyze the potential benefits and/or disadvantages in the collaboration of public schools and private agencies in providing school age child care.

For the study, the evaluators defined school age child care as follows:

> *Formal extended school day services are provided to Portland Public School children, before, during and/or after school, during holidays and vacations when working parents are unable to care for their children. These "latchkey" services are provided through contracts with parent associations and public and private agencies.*

Methods involved collecting interview and observation data on before and after school child care programs in the District, as well as information about child care policy and administration. The study focused on six district-supported pilot latchkey programs. These pilot programs were a collaboration between the school district and private, not-for-profit child care organizations. The study was organized to answer the following questions:

- What is the current status of school age child care programs in the District?
- What program, policy, and administration issues should guide District leadership in the support of existing programs and the development of new programs?

Findings, conclusions and recommendations from the study focused on: 1) a philosophy of school age child care emphasizing that children's developmental needs be met in informal learning environments which complement the school program, 2) a historical record of demonstrated need for school age child care programs in the District, 3) a synthesis of the literature on public school participation in school age child care programs, and 4) a summary of findings from site visits and telephone interviews with principals.

Following the completion of the study in 1986, the findings were presented to the Board of Education and Superintendent. The evaluators made three recommendations based on the findings of the study:

1. An administrator for school age child care should be identified to direct short term operation and long term planning of district child care programs.
2. A Task Force should be created to develop policy and program guidelines for school age child care, conduct a districtwide needs assessment, and develop a coordinated plan for school age child care.
3. The district should continue to monitor all child care programs conducted on district premises. Expansion of programs should be based on results of the needs assessment and findings of the school age child care task force.

Applications of the Study

The School Age Child Care Report was used by the School Board, Superintendent, Principals, a District Committee on School Age Child Care, the City of Portland/Portland Public Schools Liaison Office, and the Oregon State Commission on Child Care.

The Superintendent and School Board reviewed the evaluation information as evidence of the effectiveness of district latchkey programs. In response to the first evaluation recommendation, the Superintendent appointed a coordinator to direct planning and operations for District school age child care programs. Following another recommendation, the coordinator established a School Age Child Care (SACC) Committee to develop a policy statement and define the District's role in school age child care.

The evaluation also recommended that a committee develop a coordinated plan for school age child care in the District. Representatives from several public and private latchkey providers, the Portland City/Schools Liaison Office, and interdepartmental district staff met during the fall of 1986 to collaborate on the development of the policy statement on child care. The Board of Education approved the policy statement in January 1987.

Table 1 outlines the timeline of activities related to use of the evaluation. The School Age Child Care Report was published in April 1986 with subsequent oral and written summaries presented to the Superintendent, School Board, and interested community organizations. The study was well-received by the Superintendent and Board of Education. Early on, the District leadership implemented recommendations from the evaluation and established a process for long-range planning on the issue. The evaluators found utilization to be an ongoing process with planning, technical assistance, and consultation continuing periodically with district leadership about school age child care.

The first finding of the evaluation -- the need to appoint a program administrator -- was accepted and implemented by the Superintendent and School Board. While this action is a low level of utilization, in this case, it was a key factor in the continued use of the evaluation. The newly appointed school age child care administrator valued the evaluation process and was responsive to other evaluation findings.

Table 1

Timeline of Evaluation Utilization Activities

TIME	ACTIVITY
Apr 1986	Evaluation Report on School Age Child Care published.
May 1986	Evaluation findings presented to the School Board and Superintendent.
Jul 1986	Superintendent appoints District Coordinator of School Age Child Care.
Jul 1986	Oregon Commission on Child Care hears testimony by District coordinator and evaluator on school age child care in Portland.
Aug 1986	School Board charges the Coordinator with development of proposed District policy and guidelines for School Age Child Care (SACC). The Superintendent requests a report to the Board within six months.
Sep 1986	SACC Committee identified; evaluator serves on the Committee.
Oct 1986	SACC Committee begins development of policy for school age child care.
Nov 1986	Meetings with representatives of child care community to coordinate District efforts with city and state activities.
Nov 1986	School age child care policy statement proposed by SACC Committee.
Dec 1986	SACC Committee report presented to the School Board and Superintendent.
Jan 1987	School Board adopts policy statement for School Age Child Care in Portland.
1988	School Board adopts NAEYC Developmentally Appropriate Practices for Early Childhood

The second finding -- the need to develop a coordinated District policy -- was used by the School Board to support a review of long-range plans for District-supported child care. The evaluators recommended that a districtwide needs assessment be conducted, but this proposal was put on hold because of other priorities. While the evaluators expressed some limitations to the study, the findings were well accepted by the District and child care program representatives. This policy development recommendation was fully implemented by the new coordinator for school age child care. She used the report to guide the planning, organization, and management of the school age child care programs during a transition phase. The coordinator established a broad-based School Age Child Care (SACC) committee to further study the report recommendations. Ultimately, the committee used the evaluation to answer questions regarding facilities, principals' roles and responsibilities, external contractors, costs, impact on volunteer programs, and a number of other issues.

The third finding of the evaluation -- the need to monitor all child care programs conducted on district premises -- did not receive systematic approval or action.

Interpretations of Evaluation Use

For evaluation utilization to succeed, it must be incorporated as a part of the decision making process. While the primary purpose of evaluation information is to aid decision making, we discovered several different levels of evaluation use.

Figure 1 displays the Levels of Use (LoU) framework for evaluation utilization adapted by Pechman and King (1986) from the original model developed by Hall and Loucks. The framework identifies levels of school evaluation use among individual users and nonusers of evaluation. The Levels of Use model asks the question, "How is the individual using the innovation, in this case, evaluation information?" To answer the question, eight characteristics have been identified: nonuse, orientation, preparation, mechanical use, routine use, refinement, integration, and renewal. These typical behaviors help us understand the complex process of utilization while it is occurring. For our purposes, the model helped us understand what was happening in the district during the challenging and often difficult process of using evaluation information in policy development.

Figure 1

LEVELS OF EVALUATION UTILIZATION

(Adapted from Hall and Loucks Levels of Use of An Innovation, 1975)

LoU	
0	**NONUSE** User has little or no knowledge of evaluation utilization, no involvement with evaluator, and is doing nothing to become involved.
	Decision Point A. Takes action to learn detailed information about evaluation.
I	**ORIENTATION** User is acquiring information about evaluation use or has recently explored or is exploring its value and its demands upon the user and user system.
	Decision Point B. Decides to use evaluation by establishing a time to begin.
II	**PREPARATION** User is preparing for first use of evaluation.
	Decision Point C. Changes, if any, and use are dominated by user needs.
III	**MECHANICAL USE** User focuses most effort on the short term use of evaluation with little time for refection. Changes in use are made more to meet user needs than client needs. User is primarily engaged in a step-by-step attempt to master the tasks required to use evaluation, often resulting in disjointed and superficial use.
	Decision Point D. A routine pattern of use is established.
IV-A	**ROUTINE** Evaluation utilization is stabilized. Few, if any changes are made in ongoing use. Little preparation or thought is given to improving evaluation utilization or its consequences.
	Decision Point E. Changes use of evaluation to increase client outcomes.
IV-B	**REFINEMENT** User varies use of evaluation to increase impact on clients within the immediate sphere of influence. Variations are based on knowledge of both short and long-term consequences for clients.
	Decision Point F. Initiates changes with input/coordination of colleagues.
V	**INTEGRATION** User combines own efforts to use evaluation with related activities of colleagues to achieve a collective impact on clients within their common sphere of influence.
	Decision Point G. Explores alternatives to evaluation presently in use.
VI	**RENEWAL** User reevaluates the quality of evaluation, seeks major modifications or alternatives to current evaluation to increase impact on clients, examines new developments in the field, and explores new goals for self and the system.

The evaluation study provided the Superintendent with comprehensive information enabling him to develop a responsible policy for school age child care. As the Superintendent stated, "the report guided us in a unique collaboration with our local community to provide child care services to children in Portland. The report answered some of my questions, posed some new ones, and gave me a framework for policy development."

The report was useful to the Board of Education as they reviewed district practices and considered policy options for school age child care. It served as a reference for the Board of Education during two presentations and, as a result, they assigned a coordinator to look further into the issue and make recommendations for a districtwide policy. The Board Chair stated that "the thoroughness, depth of coverage, and clear-cut, easy-to-understand recommendations made the report very useful. It was well designed and well written." These and other related discussions resulted in the Districts' first policy statement on school age child care.

Community groups found the report particularly useful because it addressed the key questions for all interested parties--parents, community, care-givers, and schools. They reported that the thorough documentation of issues from each of these perspectives led the Board of Education to clarify its commitment to child care through policy decisions which supported an active partnership with the community to meet the before and after school needs of children and parents.

At the state level, the evaluation was recognized as the only written, documented study on after school child care in Oregon. The Oregon Commission on Child Care used the report extensively to analyze the viability of after-school programs. Based on the report's findings, the Commission was able to substantiate the need for similar model programs statewide. A key outcome of the evaluation utilization by this organization was the filing of a bill in the 1987 Oregon legislature that would require all Oregon school districts to allow after school programs on the same basis as the Portland model. While the bill was not enacted, the Commission was able to obtain support from major school lobby groups which had previously opposed after school use of school buildings. The Commission found two aspects of the evaluation particularly helpful: 1) the evaluation team conducted in-depth field work that a statewide commission could not have accomplished, e.g., interviewing program directors and observing program activities, and 2) the evaluation described different "types" of programs giving readers a feel for the possibilities and different approaches to school age child care. Thus, while the evaluation was originally intended to serve only Portland Public Schools, it went beyond this scope and its results were extrapolated to propose solutions to statewide problems.

Factors Contributing to Evaluation Utilization

The School Age Child Care evaluation was unique from the outset. A team problem-solving approach was used by the evaluators. The initial request from the Superintendent and School Board encompassed many concerns about child care. In order to clarify these concerns, the team discussed and outlined the key issues and problems. After coordinating the evaluation questions with District leaders, the evaluators selected the methodology and for the study. Through interviews, observations, and a review of other public school-supported latchkey programs, the evaluators outlined the history of school age child care in Portland, reported the status of current latchkey contractors, and studied the pilot school age child care programs in the district.

During the data collection phase, we conducted site visits to the six pilot programs. Data were gathered from structured interviews with the program directors and observations of the pilot child care programs. Again in this phase, the team met weekly to debrief the preliminary findings from the site visits and discuss data analysis. At this point, two members of the evaluation team agreed to collaborate on the final report.

The evaluation report was published and disseminated to appropriate users. At this point, many evaluators step away from the evaluation facilitation process feeling their role is finished. We had learned much from the initial exchanges with District leaders in clarifying and refining the key issues posed by the School Board. With this background, we decided to facilitate the evaluation use process and help decision makers achieve an operational policy and plan for school age child care. To accomplish this, ongoing discussions between the new SACC coordinator, committees, and evaluator would be required. We aimed to support of the "acquisition" of evaluation information meaning how users become informed about the evaluation, and their "utilization" meaning how the evaluation information influences the thinking, decision making, and actions of leaders. Beyond this ongoing support, it was the establishment and maintenance of "expert networks" that helped us develop a multi-faceted child care perspective responsive to the needs of different district audiences. These networks consisted of representatives from internal and external child care programs, pilot sites, and city and state child care agencies.

Timing was also a factor contributing to use of the evaluation. In a stroke of good fortune for the utilization process, 1986 was designated the Year of the Child by Congress and May was designated as local child care awareness month. Thus, the report was very timely and this created a receptive environment which certainly increased the local and regional impact of the evaluation.

In summary, the unique factors contributing to this evaluation utilization included:

- Responsive leadership - The Superintendent and Board of Education responded to the community needs for school age child care programs
- The school district is a "self-evaluating organization"
- Close coordination between the Research and Evaluation Department and the stated request for SACC information from the School Board and Superintendent
- Timely, efficient use of the evaluation's recommendations by involved audiences
- Representation on the SACC Committee by persons from various educational disciplines, departments, and public and private community agencies
- Wide interest, attention and recognition of the importance of latchkey issues by community, city, state and national organizations
- Variety of presentation styles to report evaluation findings, including written report, oral summaries to different audiences, and graphic display on school bulletin boards.
- Focused reporting on school age child care programs, policy, and administration issues. The focus recognized the need for developing a philosophy and guidelines.

Conclusion

Most users agree the most important use of the findings has been to alert the District and the community to the fact that school age child care is a critical issue in Portland. The timing of the evaluation also was a major factor in contributing to its utilization. The report was released just a week before local child care awareness month. During this time, the Portland City Club held a forum on child care and school age child care was targeted as a key issue with a reference to the Portland evaluation. This helped propel the report to wider distribution and more regional influence than it might have had at another time.

Does the evaluation continue to impact the decisions and directions of District leaders? Some users say they continue to see impact from the study today, two years after the evaluation report; others feel that we haven't gone far enough with the issue and that the most potentially influential recommendations did not receive adequate utilization and support. These users recognize that major improvements could be made in the acquisition and utilization of evaluation information and that this improvement would constitute important staff development.

Efforts to improve the way individuals use evaluation have resulted in direct use of the report, as well as outgrowths from related activities. Direct results of the evaluation included: 1) the appointment of an administrator for school age child care in the district, 2) the formation of a school age child care steering committee to study the issue, and 3) the development of a written policy statement for school age child care for the district. This statement guides the operation and management of the program. Outgrowths of the evaluation utilization process include the initiation of a study of early childhood education and its effects on later achievement of children in third grade. This study, which also impacts the school age child care population, recommended that the definition of early childhood be extended to include pre-kindergarten through grade 3 and that Board of Education adopt the National Association for the Education of Young Children's (NAEYC) guidelines for *Developmentally Appropriate Practices in Early Childhood Programs Serving Children from Birth through Eight*. These recommendations were adopted, in principle, by the District in 1988.

The topic of evaluation utilization is an important, yet complex problem. It may be that educators do not make the investments in evaluation utilization that are commensurate with the talk about the importance of databased decision making. Acquisition and utilization of evaluation take sustained effort at all levels. It is important that users see the process of utilization as a glass half-full, rather than as a glass half-empty; utilization of evaluation has promise and utility instead of only limitations. Moreover, educators will typically need sustained training and support in the utilization process to strengthen their understanding of evaluation. This training needs to move beyond the knowledge-of stage and must include practice, feedback, demonstrations, and coaching to facilitate the transfer and incorporation of evaluation information into the individual's own knowledge utilization process.

Evaluators can play a central role in establishing and maintaining the information acquisition and use cycle. Fortunately, we have a number of techniques which add to our arsenal of tactics to generate not only high utilization of evaluations, but also better information use. Looking ahead, new techniques, such as informal presentation of results and systematic communications with target audiences, will produce favorable evaluation utilization results. Hopefully, creative methods of fostering evaluation utilization will be continually tested, allowing us to better cope with the problems of information dissemination in the future.

BIBLIOGRAPHY

Alkin, M. C. & Associates. (1985). *A Guide for Evaluation Decision Makers*. Beverly Hills, CA: Sage Publications.

Alkin, M. C., & Stecher, B. (1983). Evaluation in context: Information use in elementary school decision making. *Studies in Educational Evaluation*. Los Angeles, CA: Pergamon Press, *9*, 23-32.

Alkin, M. C. & Coyle, K. (1988, April). *Thoughts on Evaluation Misutilization*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Bickel, W. E. & Cooley, W. W. (1986). *Decision-Oriented Educational Research*. Boston, MA: Kluwer Academic Publishers.

Braskamp, L. & Brown, R. (1980, November). New Directions for Program Evaluation: *Utilization of Evaluative Information*, Volume 5, San Francisco, CA: Jossey-Bass.

Glasman, N. S. (1986). *Evaluation-Based Leadership*. Albany, NY: State Univ. of N.Y.

Glasman, N. S. (1988). *Evaluation in Decision Making*. Boston, MA: Kluwer Academic Publishers.

Hall, G. E., Loucks, S. F., Rutherford, W. L. & Newlove, B. W. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption. *Journal of Teacher Education*, *26*,(1), 52-56.

King, J. A., & Pechman, E. M. (1984). Pinning a wave to the shore: Conceptualizing evaluation use in school systems. *Educational Evaluation and Policy Analysis*, Fall *6*(3), 241-251.

Moilanen, C. S. & Mitchell, S. J. (1986, April). *1985-86 Evaluation Report: School Age Child Care*. Portland, OR: Portland Public Schools.

Newman, D. L., et. al. (1980, November). "Communication Theory and Utilization of Evaluation," in Braskamp, L. and Brown, R. (Eds.), New Directions for Program Evaluation: *Utilization of Evaluative Information*, *5*, San Francisco, CA: Jossey-Bass.

Pechman, E. M., & King, J. A. (1986, April). *Analyzing the school evaluation use process to make evaluation worth the effort*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Ripley, W. K. (1985). "Medium of Presentation: Does it Make a Difference in the Reception of Evaluation Information," *Educational Evaluation and Policy Analysis*, *7*, (4), 415-426.

Paper

STANDARDIZED TESTS AT THE EARLY CHILDHOOD LEVEL: DO THEY TELL US THE TRUTH ABOUT CHILDREN?

Standardized test scores are used more often as accountability measures in even the lowest grades. Elly Pardo and Joanne Russell report the results of their investigation into differences found between a set of direct measures and a standardized achievement measure for a group of early childhood students. Their study has implications for the valid use of a standardized achievement measure for children at such a very young age.

STANDARDIZED TESTS AT THE PRE-SCHOOL LEVEL: DO THEY TELL US THE TRUTH ABOUT CHILDREN?

Elly Pardo
Joanne Collins Russell
Boston Public Schools

Viewpoints expressed in this document are solely those of the authors.

INTRODUCTION

In this paper we describe a small, pilot study investigating whether a standardized achievement test, which measures discrete items of knowledge for specific content areas, can successfully capture the abilities of young children acquiring skills through an integrated, discovery-oriented approach to learning. We suggest that because standardized measures retrieve information in ways that do not parallel the developmental learning processes of young children, they may be better indicators of good test-taking skills than of actual knowledge mastery. Evidence for our contentions comes from psychological learning theory (Piaget, 1965, 1970), observational data on the classroom of eight, first-grade students enrolled in a developmental education Early Learning Center (ELC), Metropolitan Achievement Test (MAT6) scores in reading for the eight observed students, analyses of student products, and follow-up data on student classroom performance.

Children attending the Early Learning Center in question were participating in a model developmental education program for both the Boston Public Schools and the State of Massachusetts. Consequently, a number of school-department administrators considered it important to conduct follow-up research on ELC first graders who had performed poorly on the MAT6 Reading test but who had demonstrated school academic skills that enabled them to be promoted to the second grade.

For the Boston Public Schools, poor performance on the Metropolitan Achievement Test is indicated by a score that falls at or below the 40th percentile. This percentile rank is highly important within the school department for three reasons: First, it is the point at which students are identified as being at educational risk -- that is, they are labeled as potentially in danger of failing school; second, it is the point which determines pupil eligibility for Chapter 1 remediation services;[1] and third, it is the point which flags a school, its instructional programs, its teachers, and its administrators as not having met the school system's academic standards.

The overarching purpose of a study that contrasts test performance with school achievement is to determine the level of compatibility between a standardized achievement measure used to evaluate student academic success and other indicators of student academic progress. The study described here not only examines this question but also considers a related concern, which is the extent to which a standardized test score can reliably account for a pupil's actual and potential academic attainment.

---

[1] Chapter 1 is a federal entitlement program for low achieving students.

The discussion that follows provides descriptive information on the first-grade learning environment of follow-up pupils as well as comparative data on student test scores and teacher evaluations of pupil academic performance. Analyses are presented within both a qualitative and quantitative framework and address the mismatch that has occurred between student achievement and the method used to assess instructional progress.

THE INSTRUCTIONAL CONTEXT

During the 1987-1988 school year, the students in question attended a first-grade classroom in a full-day developmental education program having before- and after-school daycare services. The philosophical assumption underlying instruction in a developmental model is that children learn best when they acquire information at their own pace through direct, hands-on experiences (NAEYC, 1988). These experiences, often individualized, are thought to allow preschoolers sufficient room for exploration, creativity, and the formation of good critical-thinking and reasoning skills that are appropriate to their developmental (cognitive) age.

Intrinsic to a developmental model is an instructional process through which child and teacher work together to plan and carry out the day's activities. Because it is widely accepted that young children acquire information holistically and not as isolated and discrete units (Piaget and Inhelder, 1969), the developmental process is based upon an integrated model of learning. This model presents activities as ongoing and interrelated (Holdaway, 1985). Themes thus become extremely important vehicles for linking various skills areas. In this way, the theme of "Beginnings" might include tasks such as planting a garden or squeezing lemons to make lemonade. These tasks, in turn, might teach a range of skills such as oral language development, counting and measuring, fine-motor coordination, and social cooperation.

Instruction in the first-grade classroom under study adhered closely to the developmental framework described above. Nevertheless, what distinguished learning in this classroom from many other first-grade classrooms was the noticeable lack of workbook and worksheet activities. Paper and pencil tasks served primarily for children to record their own discoveries and to comment on experiences by means of an inventive spelling technique.[2]

---

[2] The philosophical assumption underlying the theory of inventive spelling is that children learn to write best by developing their own guidelines of spelling and grammar. In time, as these children learn more about the adult rules of the language, their child rules will be replaced by adult forms.

Journal writing was therefore an extremely important component of the curriculum; it occurred daily and encouraged children to synthesize information, to reflect critically on their experiences from multiple points of view, to relate classroom activities to one another, and to organize their ideas into a coherent text. In the *Discussion of Data* we elaborate on the relationship of the reading-writing connection to student performance outcomes.

THE THEORETICAL CONTEXT

Current research on child learning has confirmed what researchers such as Jean Piaget (1954, 1965) had proposed much earlier - that children construct their notions of the world by physically interacting with it and not by passively observing or reacting to things and events in their environment. In fact, even around age six, when youngsters acquire the capacity to think symbolically (they can use words and numbers to represent objects and relations), their thoughts are still tied to concrete reference points. As a result, even though a child of six or seven does not have to touch or move things in order to hold them in thought, his or her representation of objects and events and the relations among them still must derive from his or her own experiences with things and situations in the world. (cf. Brown, Collins, Duguid, 1989.)

We argue here that because cognitive growth in young children occurs extremely rapidly and is subject to considerable variability within generally-recognized sequences of development, it is often difficult to obtain reliable and valid results from standardized achievement tests (cf. Cohen, 1988; Meisels, 1984). Particularly when a test is structured to tap skills mastery through an abstract, decontextualized format, a mismatch occurs between the way young children naturally internalize information and the unnatural method chosen for them to communicate what they know.

If learning takes place primarily through workbooks and worksheets, students are acquiring information by rote. They may be finding out about right and wrong answers but they are not discovering how to conceptualize and analyze problems. In fact, even though students may be 'test-ready'-- that is, they may have mastered good test-taking skills, they may not have developed the theoretical constructs necessary for them to see relationships among ideas or to apply information to a variety of learning contexts. No one can deny that youngsters are excellent imitators and that imitation plays an important role in their overall physical and intellectual growth. Nevertheless, when imitation is taught at the expense of experiential learning, complex reasoning and problem-solving skills that are critical for higher-order thinking are bypassed.

Given the above arguments, many researchers (Madaus, 1983, 1988; Meisels, S.J., 1986) and child advocacy groups (National Center for Fair and Open Testing, 1987; National Association for the Education of Young Children, 1988) have advanced the contention that tests administered on a one-time basis cannot independently account for student progress or the overall success of a given educational program. They have proposed that pupil outcomes be judged in light of a range of variables that address achievement from both a qualitative and quantitative perspective. Generally, researchers concerned with the appropriate assessment of young children have proposed that interpretations of test outcomes be made in light of three factors: the degree of fit between a test and the instructional program it measures; the cultural and language preferences of students; and the intended use of the test -- that is, how results will impact a student's educational program.

METHODS

SUBJECTS

Three males and five females participated in the follow-up study of children identified as being at risk of school failure. The children reflected the cultural and socioeconomic diversity of the Boston Public Schools: four were Black, two were Hispanic, one was Asian, and one was White; five qualified for a free school lunch. At the time they were assessed on the Metropolitan Achievement Test (MAT6), the youngsters ranged in age from 75 to 85 months; four were below the age of seven and four were seven or slightly older. All subjects were judged by their teacher to be dominant speakers of English even though two heard a language other than English in the home.

As previously mentioned, subject selection was based on a child having tested poorly on the MAT6 (at or below the 40th percentile). The percentile ranks for the children in question fell between the 5th and 37th rank.

SECOND-GRADE TEACHERS OF STUDY STUDENTS

Follow-up data on the reading/language arts school performance of observed students were obtained from their second-grade teachers. Overall, these individuals had an average of 23 years of instructional experience, with 13 of these years in a second-grade classroom. Four of the teachers said they primarily used traditional instructional techniques; three said that they used a balanced blend of traditional and developmental instructional techniques; and one indicated using more developmental than traditional teaching techniques.

In February of 1989, structured telephone interviews were conducted with each of the second-grade teachers of follow-up students. The purpose of these interviews was to determine the level of success that the students were experiencing in their reading/ language arts school program.

RATERS OF STUDENT WRITING SAMPLES

Because follow-up students had moved into more traditional learning environments than the one in which they had participated in first grade, we believed it would be informative to choose teachers for writing sample evaluations who were using instructional methods that articulated well with the format of the Metropolitan Achievement Test. In other words, we were looking for individuals who based a large part of their classroom instruction on workbook and worksheet learning. Two additional criteria for the selection of teachers were that: (1) they be assigned to schools having students with similar racial and socioeconomic characteristics as those of Early Learning Center students, and (2) overall, the 1987 - 1988 pupils of these teachers had attained average or better than average scores on the MAT6 Reading Test. The intent of this rater-selection process was to give us an indication as to whether the first-grade academic standards set for follow-up pupils were on a par with standards set by teachers of students learning successfully through more traditional methods.

Two teachers from different elementary schools were selected to conduct an evaluation of the writing samples of observed students. These individuals had an average of 24 years teaching experience, with 15 of these years in a first-grade classroom. The teachers were trained to rate transcripts of student journals independently. On four practice transcripts, their judgements for all items on the writing-assessment tool coincided 78% of the time. On all but one of the individual items, their judgements coincided either 75% or 100% of the time. (See Appendix A for inter-rater reliability coefficients.)

CLASSROOM OBSERVATION DATA

To determine how instruction was defined and administered, trained observers collected 12 hours of classroom data during April and May of 1988. All observations were conducted between 9:00 A.M. and 12:00 P.M. Data from visitations were quantified, and results indicated that 53% of the instructional activities focused on language development and literacy; that these activities occurred in a balanced mix of whole class, small group (fewer than 8 students), and individualized learning environments; and that nearly 40% of the activities were initiated by the children themselves. Quantitative findings thus show that the students in question had considerable exposure to literacy activities during the school day and that a good number of these activities were defined by the children themselves.

DATA-COLLECTION INSTRUMENTS

Several indicators of student academic achievement were used to assess skills mastery. A general description of these indicators appears below with more specific descriptions appearing in the discussion of outcome data.

Metropolitan Achievement Tests (MAT6)

The Metropolitan Achievement Tests are designed to measure student academic achievement in major content areas of the school curriculum. The tests evaluate skills appearing in leading textbook series, state curriculum guidelines, and school-system system syllabuses. Scores and corresponding percentile ranks provide general information about a student's performance within a given discipline relative to the performance of other students at the same grade level. Comparison norms are established at the national level.

Pupils in the Boston Public Schools received the Reading and Math Metropolitan Tests in early May of 1988. Outcomes for only the Reading test are considered in this study.

Curriculum Referenced Tests (CRT's)

The Boston Public Schools' Curriculum-Referenced Tests (CRT's) are in-house, school-department instruments developed to articulate with the system's curriculum objectives. Each item on a CRT corresponds to an objective, such as 'listening skills,' and clusters of test items correspond to a curriculum strand such as 'phonics' or 'word recognition.' The CRT's have been highly useful to the Boston Public Schools for several reasons: First, they afford teachers an important means for monitoring the academic progress of students against a school-department criterion; second, they provide a diagnostic tool for assessing student achievement that complements the school department's norm-referenced (MAT6) testing program; and third, by systematically capturing the same kinds of data for all grade levels, the CRT's allow for student comparisons to be made across schools within the same district and at a systemwide level.[3]

Curriculum Referenced Tests were administered to the entire student body in late June of 1988. First graders received CRT's in

---

[3] The Boston Public Schools contains five school districts.

Reading/Language Arts, Writing, Math, and Science. For the purpose of this analysis, outcome data for the Reading/Language Arts and Writing CRT's will be considered.

Student Follow-up Survey

Using a telephone interview questionnaire, teachers were asked to rate specific aspects of a follow-up pupil's grammar, vocabulary, comprehension, and writing skills as compared to those of other students in the class. Teachers were also asked to evaluate a follow-up student's progress in his or her basal reading program.[4] Additionally, they were requested to indicate their instructional style from a list of styles ranging from highly developmental to highly traditional. (See Appendix B for a copy of the Student Follow-up Survey and an explanation of the rating categories comprising the instrument.

Student Writing Sample Evaluation

Earlier we explained that journal writing was an activity of primary importance in the classroom of the students under observation. This activity is one which best exemplifies a whole language approach to the teaching of literacy. Whole-language teaching integrates reading, writing, and speaking into a single cognitive framework, its purpose being to promote holistic learning by interrelating different ways of encoding meaningful relationships (cf. Holdaway, 1979). Because reading and writing are often treated as alternative means of expressing the same cognitive processes (Kucer, 1987; Wittrock, 1983), we considered writing to be a important indicator of a pupil's reading/language arts competence.

Consequently, in January of 1989, the first-grade teacher of follow-up students transcribed the texts from journals the youngsters had produced during April of 1988 -- the month prior to their being assessed in reading on the Metropolitan Achievement test. The material in the journals covered both assigned and free-choice writing. Assigned topics related to classroom projects such as the hatching of chicks, science experiments, letter-writing, and imaginary thinking. Free-choice topics generally described experiences the children had had outside of school or within their classroom. Each journal had approximately fifteen entries.

---

[4] As of the 1988 - 1989 school year, all Boston Public School's students in grades K through 8 are required to receive reading/ language arts instruction from a single basal series. Comparisons of their basal reading progress are thus based upon the same instructional program.

Writing samples were transcribed exactly as they had been entered into a given journal, the context being provided for each individual entry. To facilitate the evaluation of written material by independent raters, specific transcription symbols flagged general grammatical or stylistic features of a pupil's writing that would be considered non-conventional within a school context. (See Appendix C for a summary of transcription conventions, and Appendix D for samples of transcribed texts.)

Instructions for the evaluation of student transcripts were as follows:

> *The rating categories below refer to aspects of a first-grade student's writing skill that determine his or her ability to compose clear and effective expository prose. You have seen that specific transcription symbols were used to flag non-conventional features of a student's writing. When rating each transcript, these 'non-conventionalisms' should not be considered as errors but rather as deviations from an adult standard. The adult standard used here is one that is appropriate to school settings and does not refer to standards appropriate to other dialects of English. As you know, many of the language deviations you have read in the child texts are normal for a first-grader; therefore, from a developmental or maturational point of view, they cannot be treated as 'mistakes.' Bearing this point in mind, please rate each transcript according to what you regard as 'extremely high,' 'average,' 'below average,' and 'extremely low' writing performance for a typical first grader.*

The four rating classifications mentioned above corresponded to the letters 'A,' 'B,' 'C,' and 'D,' respectively, and could be used with pluses and minuses to modify the value of any given classification. Each child's writing sample was judged according to the following criteria: (1) overall writing ability; (2) spelling and punctuation; (3) grammatical structure; (4) grammatical usage; (5) grammatical complexity; (6) descriptive modification; (7) thematic cohesion, and (8) idea development. (See Appendix E for a copy of the rating instrument.

ANALYSES OF DATA

Ratings from the teacher follow-up survey and student writing samples were entered into computer files. Pupil scores on the Metropolitan Achievement Test in Reading, and Reading/Language Arts and Writing scores for the Curriculum Referenced Tests were also entered into computer files. The files were then analyzed using the statistical package SPSS PC+. Analyses focused on the total-group performance of subjects, student performance by age (older than seven or younger than seven) and individual student performance. Findings showed no meaningful differences for the 'age' variable and consequently, will not be discussed in this paper. Summary data were also scrutinized to locate performance differences by race and by socioeconomic status. These differences also were not found. Finally, the data were examined to determine whether second-grade teacher ratings of pupils' classroom performance were associated at all with a teacher's instructional style. The data showed no evidence of this association.

DISCUSSION OF DATA

STUDENT FOLLOW-UP SURVEY

Teacher responses to telephone questionnaire items are summarized for the total student group in Table 1. Categories of information parallel those used to assess children on the reading survey of the Metropolitan Achievement Test. For the general areas of grammar and composition, teachers rated a child on his or her: *vocabulary skills*, knowledge of word meanings and recognition of words in context; *phonics skills*, knowledge of beginning, medial, and final sounds; *sentence-completion skills*, an ability to choose the correct word, from a list of words, that completes the meaning for a sequence of text; *literal comprehension skills*, an ability to identify the main idea and details of a story; and *evaluative comprehension skills*, an ability to predict story outcomes and logical conclusions, and to recognize the components of a story such as the 'saddest,' 'happiest,' or 'scariest' part. Teachers judged student writing skills on the basis of two criteria: *overall grammaticality*, that is, a child's use of age-appropriate features for spelling and punctuation, grammatical structure, grammatical usage, and grammatical complexity; and *overall descriptive writing ability*, a child's literary skill in using descriptive modification, thematic cohesion, and idea development. (The reader is again encouraged to refer to Appendix E for a more elaborate description of the above evaluation criteria.)

TABLE 1

STUDENT FOLLOW-UP SURVEY
AVERAGE SKILL-AREA RATINGS

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

SKILL AREA	GROUP MEAN	SCORE RANGE
VOCABULARY SKILLS	6 (B+)	4 TO 8 (B-) (A)
PHONICS SKILLS	6 (B+)	4 TO 8 (B-) (A)
SENTENCE COMPLETION SKILLS	5 (B)	3 TO 8 (C+) (A)
LITERAL COMPREHENSION	6 (B+)	3 TO 7 (C+) (A-)
EVALUATION COMPREHENSION	6 (B+)	3 TO 8 (C+) (A)
GRAMMATICAL WRITING ABILITY	5 (B)	4 TO 8 (B-) (A)
DESCRIPTIVE WRITING ABILITY	5 (B)	2 TO 8 (C) (A)

Results of teacher ratings presented in Table 1 show that on the average, students were considered to be performing at a 'B' or 'B+' level in all of the specified categories. Although scores for individual items ranged from 'C' to 'A', only one subject per category was assigned a rating lower than 'B-'.

Also comments from teachers regarding the academic performance of the students in question were all favorable. One individual who characterized her instructional style as a balanced blend of both developmental and traditional features remarked, "I really notice a difference between him and the other children. He has very good work habits and completes work on time. This may be a result of what they did with him last year. He really enjoys reading, and I know he gets support at home and that helps."[5]

---

[5] Parent involvement in their child's literacy process is an intrinsic component of the developmental education model at the Early Learning Center in question.

Another teacher with similar self-defined instructional characteristics praised a follow-up pupil's strong writing ability, explaining that "Whatever they were doing (at the ELC) in the area of writing, it was successful." Regarding three of the students in her classroom, another far more traditional teacher remarked, "I was surprised when I saw their 'MET' scores compared to the performance they are giving me. That MET test is not an indication of what a child can do!" Nevertheless, in terms of conventional school behaviors, the same teacher stated, "The kids are sooo different from any other second grader I have ever had before. They are beautiful readers, but they don't know how to sit in chairs, or they sit with their feet on chairs, or they like to sit on the floor. The three are great students...but they're into 'do your own thing.' I can't have that."

Table 2 presents a breakdown for the ratings of follow-up students' grammar and composition ability presented in Table 1. Again it should be pointed out here that the content of student-assessment categories was guided by the content of subtest areas evaluated on the Metropolitan Achievement Test. Data comparing teacher ratings of student abilities with student subtest percentile ranks for the reading survey of the (MAT6) show marked discrepancies between the assessment of a child's skills level by his or her teacher and a test's assessment of his or her skills level. For example, *vocabularly* ratings assigned by teachers are composite scores for items measuring a student's knowledge of word meanings, an ability to recognize words in context, and an ability to identify the correct word that completes the meaning for a sequence of text. If these ratings are contrasted with the percentile ranks obtained by follow-up students on the MAT6 vocabularly subtest, which also assesses the previously-mentioned skills, we notice that the information in column 1 does not correspond to the information in column 2.

In fact, with the exception of one youngster who generally scored slightly lower than the others,[6] all follow-up students were considered by their teachers to have attained a solid average or above average academic standing in grammar (vocabulary and word-recognition skills) and comprehension. The discrepancy between teacher appraisals of student classroom performance and the performance of students on the MAT6 is seen even more clearly if mean teacher ratings for specific skills areas are compared with median percentile ranks for comparable subtest areas on the MAT6.

---

[6] The teacher of this child explained that her ratings were influenced by the child's lack of motivation and not by the child's lack of training.

TABLE 2

SCORE COMPARISONS
SECOND-GRADE TEACHER RATINGS WITH MAT6 SUBTEST SCORES

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

ID	VOCAB TCHR[1]	VOCAB MAT6 NAT %ILE	WORD REC[2] TCHR	WORD REC MAT6 NAT %ILE	COMP TCHR[3]	COMP MAT6 NAT %ILE
1	8 (A)	41	8 (A)	37	8 (A)	39
2	7 (A-)	22	8 (A)	47	8 (A)	46
3	7 (A-)	31	7 (A-)	37	8 (A)	28
4	4 (B-)	18	4 (B-)	42	3 (C+)	19
5	6 (B+)	18	4 (B-)	32	5 (B)	31
6	5 (B)	14	5 (B)	32	5 (B)	28
7	6 (B+)	36	8 (A)	47	5 (B)	31
8	5 (B)	10	6 (B+)	23	6 (B+)	3
	6 (B+) MEAN RATING	*20 MEDIAN %ILE*	*5 (B) MEAN RATING*	*37 MEDIAN %ILE*	*6 (B+) MEAN RATING*	*30 MEDIAN %ILE*

Note: Ratings of student performance were obtained from the second grade teacher of each target child. The content of items evaluated by teachers parallels the content of items for the appropriate subtest on the MAT6.

[1] **VOCABULARY RATINGS** ARE COMPOSITE SCORES FOR ITEMS MEASURING A STUDENT'S KNOWLEDGE OF WORD MEANINGS, ABILITY TO RECOGNIZE WORDS IN CONTEXT, AND ABILITY TO IDENTIFY THE CORRECT WORD, FROM A LIST OF WORDS THAT COMPLETES THE MEANING FOR A SEQUENCE OF TEXT.

[2] **WORD RECOGNITION RATINGS** MEASURE A STUDENT'S ABILITY TO IDENTIFY BEGINNING MEDIAL, AND FINAL SOUNDS.

[3] **COMPREHENSION RATINGS** ARE COMPOSITE SCORES FOR ITEMS MEASURING A STUDENT'S ABILITY TO IDENTIFY THE MAIN IDEA AND DETAILS OF A STORY, PREDICT STORY OUTCOMES AND LOGICAL CONCLUSIONS, AND IDENTIFY COMPONENTS OF A STORY SUCH AS THE 'SADDEST,' OR 'SCARIEST' PART.

Data in **Table 3** contrast teacher appraisals of the skills level of follow-up students in second grade with test scores measuring their skills attainment at the end of first grade. A composite teacher rating for a student's overall grammar and composition ability was compared with the student's score on the school-department's Curriculum-Referenced Test in Reading/ Language Arts and with his or her percentile rank for Reading on the MAT6. The data show that the Reading/ Language Arts CRT more accurately captures a student's academic capability -- as evidenced by his or her second grade achievement -- than does the Metropolitan Achievement Test. This is because the CRT's are based directly on curriculum guides that drive the instructional content for the Boston Public Schools. Items on the Metropolitan Achievement Test, however, are drawn from a variety of curriculum sources nationally and span several grade levels,[7] which the CRT's do not do. Also of importance is that the CRT's are administered with more flexible time limits than are the Metropolitan Achievement Tests.

The data here show that when an instrument both articulates closely with a school system's curriculum and assesses student knowledge appropriate for a given grade level, that instrument provides a more realistic account of a pupil's actual and potential school performance than does a norm-referenced measurement tool covering material beyond a pupil's assigned grade. This finding raises questions about the 'curricular validity' of the Metropolitan Achievement Test for Boston Public Schools' students. More specifically, if an important goal of the MAT6 is to use objective criteria to identify students who may be in danger of failing academically, then it is doubtful whether an instrument that measures skills taught beyond a child's actual grade level can accomplish this objective with a high degree of accuracy.

---

[7] Items on the Metropolitan Achievement Test for Grade 1 test knowledge appropriate for Grades 1, 2, and 3.

TABLE 3

INDIVIDUAL SUBJECT ANALYSIS
READING LANGUAGE ARTS

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

ID	**GRAMMAR/COMP[1] 2ND GRADE**	**CRT: % CORRECT[2] 1ST GRADE**	**MAT6 READING[3] NAT %ILE 2ND GRADE**
1	8 (A)	93	37
2	7 (A-)	80	37
3	7 (A-)	80	27
4	3 (C+)	100	20
5	7 (A-)	73	23
6	5 (B)	80	20
7	5 (B)	87	34
8	5 (B)	93	5
	6 (B+) MEAN RATING	*86 MEAN RATING*	*25 MEDIAN %ILE*

[1] **GRAMMAR AND COMPOSITION SCORE:** TEACHER RATINGS FOR VOCABULARY, PHONICS, SENTENCE-COMPLETION, LITERAL COMPREHENSION, AND EVALUATION COMPREHENSION SKILLS.

[2] **CRT READING SCORE:** PERCENT OF ITEMS SCORED CORRECTLY ON THE BOSTON PUBLIC SCHOOLS' CURRICULUM REFERENCED READING/LANGUAGE ARTS TEST. THE CRT CONTAINS ITEMS WHICH EVALUATE A STUDENT'S GRAMMAR AND COMPOSITION SKILLS.

[3] **MAT6 PERCENTILE RANKS:** BASED ON A COMPOSITE SCORE FOR THE VOCABULARY, WORD RECOGNITION, AND READING COMPREHENSION SUBTESTS.

WRITING SAMPLE EVALUATIONS

Earlier we referred to the connection between reading and writing and argued that both processes require similar cognitive operations. These operations consist of the organization of meaning through language to define, develop, classify, and conjoin ideas. Relating experience and knowledge to a text is the foundation for both reading and writing, which are active processes of composing and comprehending. More precisely stated:

> *Composing and comprehending are process oriented thinking skills which are basically interrelated. Composing...actively engages the learner in constructing meaning, in developing ideas, in relating ideas, and in expressing ideas. Comprehending...requires the learner to reconstruct the structure and meaning of ideas expressed by another writer.*
>
> (Tierney and Pearson, 1983, p. 582)

Because of this close association between reading and writing, we decided to evaluate journal material follow-up students had produced immediately prior to the assessment of their reading/language arts skills on the MAT6. We determined that a scrutiny of pupil writing samples would provide an additional measure of their reading/language arts competence at the time of testing. Results of writing evaluations by two independent raters are presented in **Table 4.**

TABLE 4

STUDENT JOURNAL ANALYSIS
AVERAGE SKILL-AREA RATINGS

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

SKILL AREA	GROUP MEAN	SCORE RANGE
OVERALL WRITING ABILITY	6 (B+)	3 TO 7 (C+) (A-)
SPELLING AND PUNCTUATION	6 (B+)	3 TO 7 (C+) (A-)
GRAMMATICAL STRUCTURE	6 (B+)	2 TO 7 (C) (A-)
GRAMMATICAL USAGE	6 (B+)	3 TO 8 (C+) (A)
GRAMMATICAL COMPLEXITY	6 (B+)	3 TO 7 (C+) (A-)
DESCRIPTIVE MODIFICATION	6 (B+)	3 TO 8 (C+) (A)
THEMATIC COHESION	6 (B+)	3 TO 7 (C+) (A)
IDEA DEVELOPMENT	6 (B+)	3 TO 8 (C+) (A)

As expected, mean ratings of the writing samples for pupils in question showed better-than average scores for all the categories specified on the evaluation instrument. Not only did the children's texts reflect their ability to use age-appropriate grammatical features, but the texts also revealed that the children had skill in the use of descriptive terms, in relating sentences to one another, and in elaborating on a topic. Most student transcripts received scores of B+ or A-, with one transcript in each category receiving a rating of C or C+.

COMPARISONS OF FOLLOW-UP AND WRITING-SAMPLE DATA

In **Table 5** we further specify judgements of student writing ability. We begin by comparing first grade teacher ratings of a child's grammatical writing ability with second-grade teacher ratings for the same ability area. Grammar scores are composite ratings, based upon individual item ratings. The category *Grammar* refers to the use of age-appropriate spelling and punctuation, word sequences, word endings, word choice, pronoun reference, and sentence types. We also examined a pupil's stylistic writing ability. For this area, we asked first- and second-grade teachers to judge a child on his or her skill in using desscriptive modification, developing ideas, and conjoining related ideas.

Cross-comparisons of teacher judgements for both the *Grammar* and *Stylistics* categories show some similarity between contrasted ratings. Evaluations of first-grade student journals are slightly higher than second-grade teacher evaluations of overall student grammatical and writing ability. It is important to note, however, that whereas writing sample evaluations are based upon a specific product, teacher evaluations are not. Score discrepancies, then, may be attributable in part to specific versus general analyses of student performance. Nevertheless, mean ratings for the total group of observed students show a high degree of consistency across comparison categories.

An additional analysis contrasted outcome scores for follow-up students on the school-department's Curriculum-Referenced Test in Writing with ratings for student journal entries. The CRT writing test elicits three narratives, each developed around a different picture stimulus. These narratives are scored by the classroom teacher against criteria measuring the creativity of a pupil's text, its spelling, capitalization and punctuation, idea development, topic unity, and general appearance. The purpose of comparing CRT writing scores with journal scores was to contrast the judgements made by study pupils' first grade teacher with judgements of their writing ability made by other first-grade teachers.

Data show moderately high levels of agreement between teacher ratings of pupil writing competence. We have argued that writing, like reading, involves establishing relationships among words, sentences, paragraphs, and texts, and that good writing skills are closely associated with good reading skills. Teacher information regarding the progress of observed students in their second-grade basal reading program provides support for this contention. On a rating scale ranging from 'poor,' to 'excellent,' the progress of two students was judged to be 'good,' that of five students, 'very good,' and that of one student, 'excellent.' Furthermore, all the students were considered to be reading at or above grade level, and some were even at the top of their class. This finding lends further credence to the argument that standardized achievement tests may not be the most appropriate measures of actual student knowledge.

TABLE 5

INDIVIDUAL SUBJECT ANALYSIS
WRITING

RATING SCALE: 8(A), 7(A), 6(B+), 5(B), 4(B-), 3(C+), 2(C), 1(C-)

ID	GRAMMAR[1] JOURNAL RATING	GRAMMAR 2ND GR TCHR	STYLISTICS[2] JOURNAL RATING	STYLISTICS 2ND GR TCHR	CRT[3] SCORE
1	6 (B+)	8 (A)	7 (A-)	8 (A)	93
2	6 (B+)	6 (B+)	8 (B+)	8 (A)	90
3	7 (A-)	6 (B+)	7 (A)	5 (B)	93
4	7 (A-)	4 (B-)	8 (A)	4 (C)	93
5	4 (B-)	5 (B)	4 (B-)	5 (B)	90
6	6 (B+)	4 (B-)	7 (A-)	4 (B-)	93
7	6 (B+)	5 (B)	6 (B+)	5 (B)	97
8	3 (C+)	5 (B)	3 (C+)	5 (B)	87
MEAN RATING	*6 (B+)*	*5 (B)*	*6 (B+)*	*5 (B)*	*92*

Note: Ratings for first-grade student journals are based upon texts written in April of 1988. The scores reported above are an average of the composite score of each of two independent raters.

[2] **GRAMMAR RATINGS OF STUDENT JOURNALS:** OBTAINED FROM ITEMS JUDGING A STUDENT'S WRITING ABILITY IN THE AREAS OF SPELLING AND PUNCTUATION, GRAMMATICAL STRUCTURE, GRAMMATIICAL USAGE, AND GRAMMATICAL COMPLEXITY.

[2] **STYLISTIC RATINGS OF STUDENT JOURNALS:** OBTAINED FROM ITEMS JUDGING A STUDENT'S WRITING ABILITY IN THE AREAS OF DESCRIPTIVE MODIFICATION, THEMATIC COHESION, AND IDEA DEVELOPMENT.

[3] **CRT WRITING SCORES:** ASSIGNED TO STUDENTS BY THEIR FIRST-GRADE TEACHER. THE SCORES MEASURE WRITING SAMPLES ON THE BASIS OF THEIR CREATIVITY, PARAGRAPH UNITY, SPELLING, CAPITALIZATION AND PUNCTUATION, AND GENERAL APPEARANCE.

CONCLUSION

In this pilot study we have looked at several measures of student achievement in the area of reading/language arts. We have shown that a standardized achievement test administered to eight first-grade pupils enrolled in a developmental education learning program did not accurately reflect either their first-grade reading/ language arts skills level or their second-grade reading/language arts school performance. We have also shown that other indicators of pupil academic attainment such as teacher evaluations of student products and internally-developed school department criterion-referenced tests more accurately mirrored student academic capability and potential.

Our discussion has also centered on the interpretation of standardized achievement test results. We have explained that in the Boston Public Schools, end-of-year outcomes on the Metropolitan Achievement Test (MAT6) are used to classify students academically. A student scoring at or below the 40th percentile in a given academic area is formally labeled as being *at risk* of failing that area during the next school year. Whereas poor performance on the MAT6 may point to severe academic shortcomings for some students, this certainly was not the case for the eight *at-risk* students on whom we did follow-up work. Even though these youngsters had Metropolitan Achievement Test scores in Reading/ Language Arts that placed them in the danger zone according to school-department standards, their literacy skills were at or above grade level according to teacher standards.

Finally, our pilot study raises questions about the articulation between test format and instructional format and test content and instructional content. Our findings suggest that a mismatch occurs when children acquiring skills through a concrete, hands-on, and exploratory approach to learning are evaluated on a test that taps knowledge through an abstract, decontextualized format. We have noted that young children are active learners who must construct and manipulate their environment in order to develop good conceptual skills, and that rote learning for the improvement of test performance does not promote strong critical-reasoning ability. Additionally, we have proposed that tests which assess skills across several grade levels may not accurately uncover a student's actual grade-level knowledge; consequently, these instruments may not have the curricular validity they are intended to have.

Relative to policy setting for the Boston Public Schools, the implications of our findings are both programmatic and fiscal. When the results of a single test are used to label children academically, youngsters may risk being assigned to a Chapter 1 remedial program primarily on the basis of a single criterion -- a reading or math score on the MAT6 that falls at or below the

40th percentile.[8] In the case of the eight children here, four have been receiving Chapter 1 remediation services in reading, but three of the four are among the top students in their class. At a yearly cost to the school department of almost $1300 for each child, this is may not be the wisest allocation of monetary resources.

Relative to the psychological well-being of pupils, those individuals identified as being at risk in a given subject area on the Metropolitan Achievement Test may run an even greater risk during the course of their instructional program. This risk is one of being labeled and treated as a low achiever. When teachers see an extremely low score on a standardized achievement test, they may be prone to make assumptions about a child that do not necessarily characterize his or her talents. For this reason, teachers must continually be apprised of the dangers of judging student abilities on the basis of a single measure of academic progress.

In fields such as psychology, linguistics, and anthropology, information from case studies on a few subjects have helped set the stage for research and analyses of large groups of children. Therefore, given the empirical data we have presented here, we encourage school systems that use test scores to classify children to conduct follow-up research on youngsters at the early childhood level, especially those youngsters identified as being at educational risk. In this way, research findings can inform school personnel and policy makers of the degree of fit between an assessment tool and student success in the program that tool is measuring. We also encourage school systems that use test scores to classify children to think about the limitations on the information we get from a child's test score, and with these limitations in mind, to consider relying on a variety of child performance indicators when making major educational policy decisions.

---

[8] We wish to qualify this statement by indicating that all children who are eligible for Chapter 1 services do not necessarily receive them. Placement in the Chapter 1 program depends primarily on how low the child's percentile rank is on the MAT6 as compared to other children in his/her school. Schools having a great many students with scores falling well below the 40th percentile rank will have a lower cut-off point for Chapter 1 placement than will schools having fewer students in the lowest percentile ranks.

REFERENCES

Brown, J.S., A. Collins & P. Duguid. (1989). Situated cognition and the culture of learning. 18(1), 32 - 42.

Cohen, S.A. (1988). Tests: Marked for life. New York: Scholastic, Inc.

Holdaway, D. (1979). Foundations of literacy. Portsmouth, N.H.: Heinemann Educational Books, Inc.

Kucer, S.B. (1987). The cognitive base of reading and writing. In J.R. Squire (Ed.). The dynamics of language learning (pp. 27-51). Urbana, Ill: ERIC Clearinghouse on Reading and Communication Skills.

Madaus, G.F. (1983, November). Test scores: What do they really mean in educational policy. Paper presented at the 1983 Convention of the New Jersey Education Association, Atlantic City, N.J.

Madaus, G.F. (1988). The influence of testing on the curriculum. In L.N. Tanner (Ed.). Critical issues in curriculum: Eighty-seventh yearbook of the national society for the study of education (pp. 83 - 121). Chicago: University of Chicago Press.

Meisels, S.J. (1986). Testing four- and five-year olds. Educational Leadership. 44, 90 - 92.

Medina, N. & D. Monty Neill. (1988, June) Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in America's public schools. Cambridge, MA: National Center For Fair and Open Testing.

NAEYC position statement on developmentally appropriate practice in the primary grades, serving 5- through 8-year-olds. In S. Bredekamp (Ed.) Developmentally appropriate practice in early childhood programs serving children from birth through age 8. (pp. 64 - 84). Washington, D.C.: National Association for the Education of Young Children.

NAEYC position statement on standardized testing of young children 3 through 8 years of age. (1988, March). Young Children, 42 - 47.

Piaget, J. (1954). The construction of reality in the child. New York: Basic Books. Piaget, J. Piaget's theory. (1970). In P.H. Mussen (Ed.), Carmichael's manual of child psychology:- Vol. 1 (pp. 703-732). New York: Wiley.

Piaget, J. & Inhelder, B. (1969). The psychology of the child. New York: Basic Books.

Tierney, R.J. & P.D. Pearson. (1983). Toward a composing model of reading. Language Arts: Reading and Writing, 60(5), 568 - 580.

Wittrock, M.C. (1983). Writing and the Teaching of Reading. Language Arts: Reading and Writing, 60(5), 600 - 606.

APPENDIX A
INTER-RATER RELIABILITY COEFFICIENTS

PERCENTAGE OF INTER-RATER AGREEMENT

ITEM 1 = 75%

ITEM 2 = 50%

ITEM 3 = 75%

ITEM 4 = 100%

ITEM 5 = 75%

ITEM 6 = 75%

ITEM 7 = 100%

ITEM 8 = 75%

APPENDIX B

STUDENT FOLLOW-UP SURVEY

STUDENT FOLLOW-UP PERFORMANCE EVALUATION
EARLY LEARNING CENTER - DISTRICT A
TELEPHONE INTERVIEW PROTOCOL

(Name of Teacher), I'm going to ask you about *(Student's Name)*'s school performance in relation to the performance of other students in his/her class. I'm particularly interested in knowing how *(Student's Name)*'s skills compare to those of his/her peers in two broad content areas: Reading/Language Arts and Writing.

Using a scale of **A, B, C, D,** where **A** indicates *'extremely high performance,'* **B**, *'average performance,'* **C**, *'below average performance,'* and **D**, *'extremely low performance,'* I'd like you to judge the level of *(Student Name)*'s abilities against those of his/her classmates. You may use pluses and minuses with each letter you assign to a particular ability area. Let's begin with Reading/Language Arts. How would you rate *(Student Name)*'s:

1. **Vocabulary Skills**

   Knowledge of word meanings and recognition of words in context.

   A_____ B_____ C_____ D_____

2. **Phonics Skills**
   Identifying beginning, medial, and final sounds.

   A_____ B_____ C_____ D_____

3. **Sentence-Completion Skills.**
   Choosing the correct word, from a list of words, that will complete the meaning for a sequence of text.

   A_____ B_____ C_____ D_____

4. **Literal Comprehension Skills**
   Identifying the main idea and details of a story.

   A_____ B_____ C_____ D_____

5. **Evaluative Comprehension Skills**
   Predicting story outcomes and logical conclusions; identifying components of a story such as the 'saddest,' 'happiest,' or 'scariest' part.

   A_____ B_____ C_____ D_____

In the area of **WRITING**, how would you rate *(Student Name)*'s:

6. **Overall grammatical writing ability**

Includes age-appropriate features for: (1) **spelling and punctuation**; (2) **grammatical structure** (appropriate word sequences, the use of correct person and tense endings on verbs, the use of appropriate plural markers on nouns); (3) **grammatical usage** (appropriate word choice and pronoun reference); and (4) **grammatical complexity** (the use of different sentence types -- simple, compound, complex).

A_____ B_____ C_____ D_____

7. **Overall descriptive writing ability**

Includes: (1) **descriptive modification** (appropriate use of adjectives and adverbial modifiers); (2) **thematic cohesion** cohesion (sentences are related to each other in a meaningful way); and (3) **idea development** (the ability to elaborate on a topic).

A_____ B_____ C_____ D_____

Now I have just a few more general follow-up questions.

8. How is *(Student)* progressing in his/her basal reading program? Would you say his/her progress is:

A. excellent B. very good C. good D. fair E. poor

Name of basal reading series_______________________________

Name of book student is using in series_______________________

Level of book in series_____________________

9. For approximately how many hours a week does *(Student Name)* engage in self-directed reading with texts other than a basal?

____

10. For approximately how many hours a week does *(Student Name)* engage in self-directed process writing?

____

11. In terms of a continuum of teaching styles where a **developmental teaching style** is characterized by hands-on, discovery-oriented, student-directed, and process-oriented learning and a **traditional teaching style** is characterized by paper and pencil, workbook, teacher-directed, and product-oriented learning, how would you characterize your instructional style? Would it be:

    A. completely developmental

    B. primarily developmental with some traditional features

    C. a balanced blend of developmental and traditional features

    D. primarily traditional with some developmental features

    E. completely traditional

APPENDIX C

TRANSCRIPTION CONVENTIONS

STUDENT TRANSCRIPTS
KEY TO TRANSCRIPTION CONVENTIONS

∧ **OMISSION OF PUNCTUATION MARKS, CAPITAL LETTERS, OR INCORRECT PUNCTUATION**, e.g., ',' for '.'

I like chicks, they are beautiful.

o **OMISSION OF A PERSON/TENSE MARKER, AN AUXILIARY VERB, OR A PLURAL MARKER ON A NOUN**

Yesterday I go to the movies; I been there many times; There are two chick in the incubator.

X **OMISSION OF A WORD OR SEQUENCE OF WORDS**

We planted and the seeds were going to grow into flowers.

〰 **UNCONVENTIONAL SPELLING OR RUN-ON WORDS**

I played a horsegame.

○ **INCORRECT WORD USAGE**, e.g., 'a' for 'an,' 'the' for 'a,' incorrect pronoun reference

Once upon a time there was the dog named Sam.

It's a nice horse, and I think he's pretty.

___ **OVERUSE OF CAPITALS/PUNCTUATION**

The book is on The table near the Chair.

∽ **REVERSE WORD ORDER**

He helped me it do.

(- - -) **UNCLEAR/GARBLED TEXT:** cannot decipher meaning

(=   ) **MEANING GLOSS:** interprets or paraphrases student's intended meaning.

Tanscript # 403

Page 3 of 5

Context	Text	Page #
Writing about chick development incubator in room.	^in The Egg There is ... a little chick^ The yolk is yellow. ^an (=And) Tamarole (=Tomorrow) The chicks will Hach^ (=hatch)	19
Free choice writing.	When I go over my cousent (=cousin's) house I will Take my Mashen (=machine) over^ ^in The Summer we got some Ice carem (=cream) an (=and) we got to go to toys Я us^ ^when we got to my cousent house we road (=rode) my bick (=bike) an (=and) we went to The sTore an (=and) x (=we) went to The faire^ ^we got a new dugary (=dungaree) cote^ (=coat) ^we had no chraing (=training) wees^ (=wheels);	20
Free choice writing.	My los (=loose) tooth is wigaly (=wiggly) an (=and) I dont^ care if it fall^ ^the tooth firy (=fairy) will give me a dola (=dollar) or a koer^ (=quarter) ^she gave me a note^	23

Transcript # 404 Page 4 of 5

Context	Text	Page #
Free Choice Writing.	My mom Sed (= said) She would by me a netedo (= Nintendo) viteeo (= video) game At the Ennd (= end) of spring. I am going to play a game. I am going to hook it on my TV.	18
Writing about school cookout.	We had a cook out out side in the play yard. We had a hot dog And Chips And iy-scren. (= ice cream) And we playd. (= played)	21
Free Choice Writing.	At the park I like to play socr (= soccer) ball outside. I am the gollee. (= goalie) I catch the ball all the time.	22

Transcript # 405 Page 3 of 5

Context	Text	Page #
Free choice writing - based on recent science lessons.	a comet is made out of rock and metal and Ice. @ Asteroid is made out of rock and Fire. a ring of a planit (=planet) is like a frisbe (=frisbee). a star is wite (=white) and a star (=is) mad (=made) ou (=out) of wite rocks.	13
Writing about self as a chick.	When I was a chick I saw some horsEs eating hay in The farm. And I saw Some geeses and ducks and hens.	14
Writing about chick development - incubator in room.	In @ egg thers (=there's) a little thing. Then it grose (=grows) Bigr (=bigger) and Bigr until they hatch out from The egg.	17

APPENDIX E
WRITING SAMPLE EVALUATION INSTRUMENT

Evaluation Instrument

STUDENT WRITING SAMPLES

Transcript #__________

The rating categories below refer to aspects of a first-grade student's writing skill that determine his or her ability to compose clear and effective expository prose. You have seen that specific transcription symbols were used to flag non-conventional features of a student's writing. When rating each transcript, these 'non conventionalisms' should not be considered as errors but rather as deviations from an adult standard. The adult standard used here is one that is ap-propriate to school settings and does not refer to standards appropriate to other dialects of English. As you know, many of the language deviations you have read in the child texts are normal for a first-grader; therefore, from a developmental or maturational point of view, they cannot be treated as 'mistakes.' Bearing this point in mind, please rate each transcript according to what you regard as 'extremely high,' 'average,' 'below average,' and 'extremely low' writing performance for a typical first-grade student.

RATING SCALE
PLEASE CIRCLE THE LETTER OF YOUR RESPONSE
To qualify your judgements, you may add pluses and minuses to your letter choice, e.g., B+, C-, etc.

A__ 'extremely high'
B__ 'average'
C__ 'below average'
D__ 'extremely low'

1. *OVERALL WRITING ABILITY*

A__ B__ C__ D__

2. *SPELLING AND PUNCTUATION*
The use of age-appropriate spelling, word segmentation, and punctuation.

A__ B__ C__ D__

RATING SCALE
PLEASE CIRCLE THE LETTER OF YOUR RESPONSE

To qualify your judgements, you may add pluses and minuses to your letter choice, e.g., B+, C+, etc.

A__ 'extremely high'
B__ 'average'
C__ 'below average'
D__ 'extremely low

3. *GRAMMATICAL STRUCTURE*

Appropriate word sequences; the use of correct person and tense endings on verbs; the use of appropriate plural markers on nouns, e.g., He a real nice person; I planted a seed, and it grow into a flower. I have two best friend.

A__ B__ C__ D__

4. *GRAMMATICAL USAGE*

Age-appropriate word choice and pronoun reference, e.g. Once upon a time there was the dog named Sam; It's a nice horse, and I think he's pretty;

A__ B__ C__ D__

5. *GRAMMATICAL COMPLEXITY*

Age-appropriate use of different sentence types, e.g., **simple**: We couldn't play outside. It was raining; **compound**: It was raining and we couldn't play outside; **complex**: sentences with subordinators such as when, because, who, which, that, where, if, so, etc.: We couldn't play outside because it was raining; I know where I'm going to plant my garden; She's the witch who's good.

A__ B__ C__ D__

6. *DESCRIPTIVE MODIFICATION*

Age-appropriate use of descriptive adjectives and adverbial modifiers, e.g., Yesterday my mother bought me bright red pants and a shiny blue shirt; as compared to: My father bought me pants and a shirt.

A__ B__ C__ D__

7. *THEMATIC COHESION*
Sentences are related to each other in a meaningful way.

A__ B__ C__ D__

8. *IDEA DEVELOPMENT*
The student's ability to elaborate on a topic.

A__ B__ C__ D__

Survey
TESTING PRACTICES AND ISSUES

Ernest Bauer agreed to supervise a survey of NATD members. The results reported here are based on 50 respondents representing school districts with a median population of 42,000 students.

TESTING PRACTICES AND ISSUES

Ernest Bauer
Oakland (MI) Public Schools

Abstract: A survey about the testing programs in 50 NATD members' mostly large school districts revealed that 78% annually test every student in grades K, 1, or 2 through 9 with a commercially-procurred test. The 3.7 million students annually take 2.6 million standardized tests and and 2.3 million locally developed tests. Commercially-procurred tests are used primarily for evaluation, accountability, and diagnosis. Locally developed tests are used primarily for diagnosis and evaluation. State mandated tests are used primarily for certification and are much more likely to occur at the high school level. The total budgets for the 50 testing departments was $8.6 million. The average testing department total expenditure divided by the total K-12 enrollment ranged from $.84 to $19.39. Most districts are using only national, school mean, and/or local norms. 84% of the respondents believe that publishers need to coordinate their standardization efforts. Most cooperate with publishers collecting test standardization data less than 50% of the time.

To help fulfill its mission of sharing practical information among test directors, the National Association of Test Directors (NATD) conducted a survey of testing issues and practices in members' public school districts in the United States during the 1988-89 school year. The survey was mailed to all 210 NATD members. About 80 members are not employed as public school testing directors and were not expected to respond. Some educational organizations have several persons who are members and only one response per organization was solicited, so the 50 public school respondents represents about a 40% response rate. Eleven persons responded who were not in local educational agencies. Their responses are not included in this report.

The Districts: NATD was originally an organization of large district testing directors. It still is. While one member reported that her district's enrollment was 3,300 students, the median was 42,000 students. Only 25% of the respondents had fewer than 27,000 students in their districts. Twenty-five percent of the respondents had more than 82,000 students in their districts. One district reported a total K-12 enrollment of 1,000,000 students. The total number of students represented by the 50 districts was 3,678,000.

Twenty-eight of the districts are organized with kindergarten (K) through 5th grade in elementary schools; 6th through 8th grades in junior high or middle schools; and grades 9 through 12 in high schools. Thirteen had a K-6; 7-8; 9-12 structure. Six were organized as K-6; 7-9; 10-12 districts.

* The author thanks Paul LeMahieu of Pittsburgh Public Schools for his assistance in constructing the survey.

The respondents tend to be persons who devote the majority of their professional time to testing matters: 15 (30%) of the respondents reported spending 100% of their professional time on testing matters. The average respondent reported spending 68% of his/her time on testing matters. The 50 respondents represented 135 professional staff members who work directly with their district testing program. They reported that their districts collectively employ forty-seven technicians, 68 clerks, 57 secretaries, and 18 other persons who work directly with their district testing programs. The individual district's office size ranged from 1 to 45 employees. Twenty-five percent employed less than 3 persons and 25% employed 6 or more persons. The median number of testing employees in the districts was 4.

The Services: Respondents were asked to check which of 9 computerized testing services their offices provided. Table 1 shows that most of the services were provided by most of the districts. The first two services (both are related to item banking) were the only ones provided by less than half of the districts. The average office provided 6.5 of the 9 services. Only 25% offered 5 or less and 25% offered 7 or more.

Number of Testing Offices Which Provide This Computerized Service	Area
20	Storing test items
16	Generating tests from stored items
35	Pre-gridding student info. on answer sheets
36	Scanning & scoring locally developed tests
38	Scanning & scoring commercially procurred tests
45	Storing test scores
42	Integrating test results into student databases
45	Producing statistical analyses
41	Preparing graphic presentations of test results

Table 1. The services offered by the testing departments in 50 large city public schools.

The testing directors were also asked on what kind of computer most of their testing work is done. Thirty-seven said on a mainframe, 3 on a minicomputer, and 6 were using microcomputers for the majority of their testing work. One said the item did not apply and 3 did not respond. It appeared that nonresponses came from offices which have the majority of their testing work done by outside vendors.

Issues and Practices: There has been a lot of publisher activity centered around different types of norms. Which are being used? If not, are testing directors planning to begin using them? Of the 8 types of norms currently available on some nationally standardized tests, only three are being used by more than 1 or 2 large district testing departments: national norms, school mean norms, and local norms. These were used by 50, 25, and 23 districts respectively. Two districts reported using large city or urban norms and low socioeconomic status (SES) norms. Only one district reported using high SES norms and rolling norms. No one reported using suburban norms. Nineteen of the testing departments use only 1 of the 9 types of norms; 12 use 2 types; 17 use 3 types; and 2 use 5 different types of norms. There is, however, considerable interest in adding new types of norms to the testing departments' repertoires. Thirteen reported that they plan to begin using rolling norms. Nine would like to add low SES norms; 8 would like to add high SES norms. Seven plan to begin using local norms; 4 plan to add large city norms; 3 plan to begin using school average norms.

Relations with Publishers: Most districts rely on commercially procurred tests for part of their testing program, so district/publisher relationships are important. Twenty-four respondents reported they are licensed to score commercially-procurred tests. Several commented that they did not believe they need to be licensed to score for their own district. Fifteen of the 24 who have licences reported they would be willing to share information about the royalties they pay.

Thirty-three use custom designed scannable documents. (This was not limited to answer sheets for commercially procurred tests and should not be construed to mean that several are printing answer sheets without permission.) Twenty-seven are using continuous form answer sheets. (That means they can be used by a computer printer to pregrid certain student information, like name, gender, grade, and student number.) There was a widespread belief that answer sheet vendors are currently "custom designing" the same answer sheet for many users. Twenty-five of 32 respondents said that they "..suspect that several other agencies use custom documents similar to yours." Twenty-one expressed an interest in knowing who those other agencies are. Twenty-six of 50 repondents said they had purchased a norms table tape from a publisher. Twenty-one said the norms table tapes were reasonably easy to use and 18 found the original version to be correct.

One important reason that the NATD was formed was to try to establish a vehicle to help inform publishers that it is very difficult for large districts to participate in every national standardization sample. Publishers report they cannot coordinate their standardization efforts without charges of collusion. Large city school districts tend to be included in every representative national sample and find it difficult to cooperate with every request. The testing directors clearly see this as an important issue. Forty-two said it was either a very important issue (22) or a somewhat important issue (20). Only 1 said they always cooperate with publishers who need students and only 1 said they always refuse. Twelve reported they cooperate 50-74% of the time; 14 25-49% of the time; and 16 reported they cooperate 1-24% of the time; so the majority cooperate with publishers less than half the time. Testing directors are aware of this and it causes concern about the representativeness of the national samples that are used for standardized tests.

Local Policy about Testing: Testing directors are keenly aware that test scores are frequently misused by the public and by school administrators. While the existence of written policy about testing programs does not assure appropriate use or good practice, it does show that concerns about testing were brought to the boards' attention. The testing directors were asked to report about the existence of written policy about their program in general and about several specific areas. Table 2 shows the distribution of their responses.

Q: Does your district have written, board adopted policies pertaining to your testing program?	24 Yes	18 No
If so, do they cover:		
Selection of tests	6 Yes	9 No
Use of test results	19 Yes	15 No
Who can be tested	12 Yes	15 No
Access to scores	18 Yes	9 No
Restrictions on analysis	1 Yes	25 No
Restrictions on dissemination	7 Yes	20 No

Table 2. The distribution of responses to items about written, board adopted policies pertaining to testing programs.

Most district's written, board approved policies cover only one, two, or three of the areas listed (6, 7, & 7 districts respectively). Only 1 district had policies which covered more than 5 of the areas.

The Testing Programs: The primary question on the survey was, "Describe the testing program for which you are administratively responsible." Respondents were asked to include norm and criterion referenced tests that are part of their programs. A set of codes was provided to represent the major purpose of each test. The following information was also requested for each test: grade level, number of students tested, the month of testing, and whether the tests were scored locally, and approximate cost.

It is difficult to briefly summarize the testing program descriptions, however, a few major features were evident. First, 29 (58%) of the testing programs included yearly testing of nearly every pupil with a standardized, commercially-procurred test. (Programs starting in grades K, 1, or 2 and incorporating every grade level through 10, 11, or 12 were included in this category.) Ten (20%) programs included students in every grade from K, 1, or 2 through grade 9 but did not include grades 10 through 12. In 78% percent of the districts (58% that test K-12 plus 20% that test K-9) every student in every grade from K, 1, or 2 through 9 is tested annually. Nine of the programs tested students at certain grade levels only. Only 2 reported that their testing program did not include standardized tests. The 3.7 million students in 50 public school districts took 2.6 million standardized tests, 740,000 state mandated tests, and 2.3 million locally developed tests.

Sixteen of the programs used combination aptitude/achievement testing in one or more grade levels. Thirty (60%) of the programs used locally developed objective referenced tests. Seventeen testing directors reported state mandated tests were part of their program.

Test Administrations: To analyze the program descriptions, each test administration (a test at a grade level at a time) was made the unit of observation. There were 617 occurances of standandized tests, 179 occurrances of state mandated tests, and 458 criterion referenced tests. The average district's testing program, then, included 12 administrations of standardized tests, 3 administrations of state mandated tests, and 9 administrations of criterion referenced tests. The major purposes of the different types of tests are shown in Table 3.

Major Purpose of the Test	Type of Test		
	Commercial	State	Local CRT
Diagnosis (instructional management)	55%	18%	69%
Placement (e.g., kindergarten readiness, adv. placement)	25	12	7%
Guidance (e.g., vocational interest)	12	3	5%
Selection (gifted or other special program)	47	2	6%
Certification (promotion or graduation requirement)	18	64	24%
Accountability (status of students)	66	42	27%
Evaluation (program evaluation)	72	21	50%
Number of Test Administration:	617	179	458

Note: Percents add to more than 100 because each test could be used for several purposes.

Table 3. The Major Purposes of the Tests Given in 50 Testing Programs.

Most of the commercially-procurred tests are being used for evaluation (72%), accountability (66%), and diagnosis (55%). Most of the state mandated tests are used for certification (64%). Most of the local criterion referenced tests are used for diagnosis (69%) and evaluation (50%). It was surprising to this author that only 12% of the tests are used for guidance as guidance counselors are frequently seen as "testing experts" in many public school districts.

Ninety-six percent of the locally developed test administrations are also scored locally. Forty-nine percent of the comercially-procurred tests are scored locally. Thiry-five percent of the state mandated test administrations are scored locally.

There is always a pull between spring and fall testing in public schools. With fall testing, the teacher who administers the test is also there to receive and use the results. With spring testing, the teacher knows what program of instruction the children have actually had for the past 7 or 8 months. April and May are the months in which the most testing occurs (260 and 290 administrations, respectively). October is the third most popular month with 187 test administrations. Somewhat surprising is the popularity of January with 139 test administrations. January has become a more popular testing month in some districts because it makes recent test data available for February or March scheduling.

September	120	January	139	May	290
October	187	February	53	June	33
November	10	March	121	July	12
December	27	April	260	August	1

Table 4. The Month in which Test Administrations Occur in 50 Large Public School Districts.

As mentioned above, most of the programs include nearly every pupil at every grade in grades K, 1, or 2-9. Table 5 shows the actual distribution of the test administrations at each grade level. Kindergarten and 12th grades are the least frequently tested grades when local districts are the testing agency. In contrast, in state mandated testing, 12th grade testing is nearly as likely as several other grades. In fact, more state mandated testing is ocurring at the senior high levels than at the elementary level and junior high levels in sharp contrast to standardized and criterion referenced testing.

	Type of Test:					
Grade Level	Commercially Procured		State Mandated		Locally Developed	
	No. of Students (in 1000's)	No. of Administrations*				
Pre-K	8	(2)	0	(0)	0	(0)
K	106	(26)	0	(0)	26	(9)
1	212	(43)	16	(5)	127	(22)
2	210	(47)	3	(1)	146	(25)
3	266	(65)	64	(25)	231	(42)
4	218	(63)	10	(4)	220	(39)
5	262	(67)	26	(7)	218	(41)
6	223	(62)	54	(21)	248	(47)
7	225	(61)	30	(8)	280	(53)
8	282	(72)	75	(23)	302	(59)
9	190	(53)	104	(21)	206	(41)
10	161	(43)	139	(27)	173	(34)
11	146	(39)	114	(20)	116	(29)
12	106	(28)	105	(18)	28	(17)
Total	2,617	(617)	739	(179)	2,318	(458)

Grand Total: 5,676,498 students tested in (1,308) test administrations.

* An administration is a test given at a grade at a time.

Table 5. The number of students tested and the number of test administrations at each grade level in 50 districtwide testing programs in the United States.

The Cost of the Testing Programs: The Testing Directors were asked to place an approximate cost on each element of their testing programs. Apparently, this is too difficult to do because none of them did so. However, 38 of them did respond to a question about the overall cost of the testing programs which they administer. The item on the survey read, "What is the total annual cost of your districtwide testing program (not including in-kind costs)?" Blanks were provided for "Personnel Costs" and

"Other Than Personnel Costs." The total for personnel costs for 34 testing programs was $3,845,782. The total for other-than-personnel-costs for 37 programs was $4,742,669, for a grand total of $8,588,451 annually. The personnel costs ranged from $8,000 to $350,000 annually with an average of $113,111. The other-than-personnel-costs ranged from $1,000 to $1,100,000 with a mean of $128,180. Total costs reported for 38 testing programs ranged from $21,000 to $1,450,000 with a mean of $226,011. 25% of the programs cost less than $90,000 and 25% of the programs cost more than $234,000. The median total program cost, not including in-kind costs, was $127,000.

Since most testing directors did not provide cost data on a per test basis, it is not possible to compute an exact test per pupil cost. However, 38 districts' total K-12 enrollment and total testing program costs were available. While the result of dividing these two numbers is not a per pupil per test expenditure, it is the total expenditure for the average student. Table 6 shows the resultant quotients. Twenty-five percent of the districts reported spending an average of $2.30 per pupil or less. Twenty-five percent spent $6.14 or more. The median average per pupil cost was $3.73. The average for the 38 district averages was $4.79.

Cost Range:	No. of Districts:				
$ .84 - 1.00	2	$4.01 - 5.00	5	$7.01 - 8.00	3
$1.01 - 2.00	4	$5.01 - 6.00	3	$8.01 - 9.00	2
$2.01 - 3.00	8	$6.01 - 7.00	2	$9.01 -10.00	1
$3.01 - 4.00	6			$10.01 -19.39	2

Note: These averages are simply the total cost of the district's testing program divided by the total K-12 enrollment.

Table 6. The average per pupil expenditure for testing in 38 large public school districts.

The testing directors were asked, "What are your major concerns about testing?" Several themes ran through the 101 responses received. The most prevalent were: (1) the appropriate use of tests with various publics; (2) the impact of externally mandated tests; (3) the technical adequacy of tests; and (4) issues of curriculum alignment.

Finally, the testing directors were asked what NATD should be doing to better serve their needs. It appeared that many testing directors would like to see the National Association of Test Directors take a more proactive role in promulgating good testing practice. The ideas of NATD serving as a clearinghouse, publisher, and/or author were the recurring themes in the 43 responses to this item. Several suggested that surveys like this one should become a part of the National Association of Test Directors' regular agenda.